# Identification of Regulatory Regions in the *Drosophila dmyc* Gene: Bioinformatics Analyses Combined with Reporter Activity Studies

**Dissertation**

zur

**Erlangung der naturwissenschaftlichen Doktorwürde**

**(Dr. sc. nat.)**

**vorgelegt der**

**Mathematisch-naturwissenschaftlichen Fakultät**

der

**Universität Zürich**

von

**Jasmine Kharazmi**

von Zürich (ZH)

**Promotionskomitee**

Prof. Dr. Hans-Peter Lipp (Vorsitz)

Prof. Dr. François Karch

Prof. Dr. Bernard Schmid

Dr. Daniel Bopp

**Zürich, 2013**

To the University of Zurich, my children Cameron and Sarah, to my mother, and in the memory of my father

Die vorliegende Dissertation wurde am 7. Dezember 2012 zur Begutachtung eingereicht.

Promotionskomittee:

Prof. Dr. Hans-Peter Lipp (Vorsitz)

Prof. Dr. Bernard Schmid

Prof. Dr. François Karch

Dr. Daniel Bopp

# Table of Contents

**Abbreviations**

| | |
|---|---|
| aa | amino acid |
| AMV | Avian Myeloblastosis Virus |
| APC | Adenomatous Polyposis Coli Tumor Suppressor |
| attB | Bacteriophage ΦC31 attachment Site B |
| attP | Bacteriophage ΦC31 attachment Site P |
| BAC | Bacterial Artificial Chromosome |
| BDGP | Berkeley *Drosophila* Genome Project |
| bHLHLZ | Basic region/helix-loop-helix/leucine zipper |
| β-Gal | β-Galactosidase cDNA from *E. coli* (see *lacZ*) |
| CSBs | Conserved Sequence Blocks |
| CTD | C-Terminal Domain |
| DGRC | *Drosophila* Genome Research Center |
| DH | DNase I Hypersensitive Site |
| DPE | Downstream Promoter Element |
| ESC | Embryonic Stem Cell |
| Inr | Initiator Element |
| iPSCs | induced Pluripotent Stem Cells |
| *lacZ* | β-Galactosidase cDNA from *E. coli* |
| LPA | Lipid Growth Factor |
| MB1 | Myc Box1 |
| MB2 | Myc Box2 |
| MBNbs | Mushroom Body Neuroblasts |
| MC29 | Myelocytomatosis Virus (a Defective Avian Retrovirus) |
| MCS | Multiple Cloning Site |
| MF | Morphogenetic Furrow |

| | |
|---|---|
| NLS | Nuclear Localization Sequence |
| NTD | N-Terminal Domain |
| ORF | Open Reading Frame |
| pC | plasmid CaSpeR4 |
| PCR | Polymerase Chain Reaction |
| RACE | Rapid Amplification of cDNA Ends |
| RT-PCR | Reverse Transcription Polymerase Chain Reaction |
| SOP | Sensory Organ Precursor Cells |
| T-ALL | T-Cell Acute Lymphatic Leukemia |
| TCF | T-cell Specific Factor |
| TFBMs | Transcription Factor Binding Motifs |
| UTR | Untranslated Region |
| WB | Western Blot |
| X-Gal | 5-bromo-4-chloro-indolyl-β-D-galactopyranoside |

**Summary**

A fundamental question about the biology of the developmentally regulated *myc* gene is how both its high expression in growing and dividing cells as well as its down regulation during differentiation is maintained. The product of the proto-oncogene *myc* is a transcription factor, a crucial regulator of growth and proliferation during animal development. As an evolutionarily conserved gene regulatory protein, Myc contains a sequence specific "basic region/helix-loop-helix/leucine zipper" (bHLHLZ) DNA binding domain and an N-terminal transactivation domain, both essential for its biological activity in cell growth control, differentiation and apoptosis.

Understanding the regulation of mammalian *c-myc* as an "immediate early gene" at the level of transcription is still a challenge. Due to the presence of only one copy of the *dmyc* gene (instead of at least three isoforms as in mouse), and well-established genetics of the model organism *Drosophila*, it is possible to study the *cis*-regulatory elements in *dmyc*.

The aim of this study was fourfold: (i) to identify *cis*-regulatory control elements in the 5'-UTR, intragenic region, and 3'-UTR responsible for the expression of the endogenous *dmyc*; (ii) to elucidate the mechanism of transcription termination; (iii) to characterize putative transcription initiation sites; (iv) to define the mechanism of *dmyc* RNA splicing. To answer these questions, three approaches were applied: 1) bioinformatics-based analyses for revealing the complexity of *dmyc* transcriptional regulation as well as for obtaining guidance for designing the experiments to test various transcriptional control elements/regions of the *dmyc* gene; 2) study of *lacZ* reporter activity under the control of the *dmyc* promoter and other potential regulatory regions; and 3) application of 5' RACE (Rapid Amplification of cDNA Ends) and subsequent sequencing of data obtained to map the start of the transcript and elucidate the mechanisms of RNA splicing. The approaches used in this study enabled us to obtain the results explained bellow and draw the following conclusions:

Alignment of a 40-kb reference DNA fragment on the X chromosome, harboring the *dmyc* gene of *D. melanogaster*, and orthologous sequences from 12 sequenced Drosophilids with the phylogenetic footprinting bioinformatics tools *EvoPrinter* and *cis*-Decoder identified several putative enhancer regions with conserved sequences, including several conserved E-boxes, which are preferred Myc binding sites of both

the CACGTG and CACTTG type. In the intron 2 region, in addition to two conserved E-boxes we identified multiple clusters of conserved sequence blocks upstream of the predicted intronic promoter. Two of the several identified clusters of conserved sequences include: (i) a repeat sequence element (ATGTTGCCA) where the core (TGTTGC) is repeated three times and (ii) a dead-ringer for the HLHm-3-2 enhancer (CGCGTGGGAAAA), within which the consensus binding site (GTGGGAA) for *suppressor of Hairy wing su(Hw)*, resides. In the large 3'-UTR region, we identified clusters of conserved sequence blocks, but no E-boxes. A search with the bioinformatics tool neural network genetic algorithm PROMOTER 2.0 detected a potential promoter region (dubbed as P0) in the far upstream region, and a search with DNASTAR Lasergene 9.1 (module: GeneQuest) identified the P1 promoter in the proximal 5'-UTR sequences, the P2 downstream promoter element (DPE) in intron 2, and three potential polyadenylation signals in the 3'-UTR.

On the basis of the computational analyses and the appearance of suitable restriction sites in the noncoding regions, different deletion constructs were made. The transgenes were injected either into the $y^1$ $w^{1118}$ or attB fly embryos to establish independent transgenic stocks. The activity of *lacZ* reporter was analyzed in third instar larval brain, imaginal discs, embryos and ovaries of the established fly lines.

In the brain *lacZ* is expressed in a number of cells mostly in proximal parts of both hemispheres and along the ventral ganglion. These cells could refer to neuronal precursor cells, which undergo a high rate of growth and cell division before differentiating. The *lacZ* activity in the cells in the front parts of the hemispheres could represent a certain class of continuously proliferating neuroblasts, namely mushroom body neuroblasts (MBNbs). In the wing disc *lacZ* is mostly expressed around the wing pouch and in the notum region. In the eye disc *lacZ* is predominantly expressed anterior and posterior of the morphogenetic furrow (MF), and in the antennal disc in the central region. In the leg disc the reporter activity is observed in concentric rings in the middle of the disc with a hole in the center. The single cells in the leg and wing discs staining for lacZ could refer to sensory organ precursor cells (SOPs) that divide actively before terminal mitosis and differentiation. In the early embryo maternal transcripts are detectable, later during germ band extension *dmyc* intensifies in the mesoderm, mid-gut, pharynx, and anal pad. In the ovary, *dmyc* activity is predominantly detected in the nurse cells and oocyte, and weakly at the tip of

germarium. The patterns observed in the tested tissues correlates with endogenous *dmyc* localization.

The expression obtained with the construct containing only intron 2 sequences indicates the existence of a downstream initiation site driving the transcription of the reporter gene. This corresponds to the finding that many of the developmentally active genes contain modular *cis*-regulatory elements in order to switch the gene on and off during different stages of development.

From the results obtained, I conclude that most, if not all of the regulatory elements required for the correct expression of *lacZ* in larval brain and discs, embryos, and adult female tissues are present within the tested 40-kb fragment of the *dmyc* locus. I further conclude the sequences including intron 1, the 5'-UTR, and approximately 100 bp upstream of the transcription start site from P1 promoter is sufficient to give reporter activity in a *dmyc*-like pattern in both ovarian oocyte, nurse cells, and in the embryo, but not in larval tissues. Finally, like many developmentally regulated genes, *dmyc* is likely to be transcribed from multiple transcription initiation units including a far upstream regulatory region, a TATA box containing proximal complex and a TATA-less downstream promoter element in conjunction with an initiator within the intron 2 region.

Analysis of the polyadenylation signals, poly (A)1, poly (A)2, and poly (A)3, at the *dmyc* 3'-end revealed that the *dmyc* gene would be predicted to produce different transcripts with shorter and longer lengths, as known from literature.

5’ RACE analysis of the *dmyc* cDNA ends mapped the transcription start site at the P1 promoter 18 base pairs upstream of the start of the known EST GM01143 within the 5’-UTR, an A residue being transcribed as the first base. The data show that the first TATA box, previously computationally predicted, is utilized to generate *dmyc* full-length mRNA. The largest transcript contains all three exons generated after the removal of the introns by regulated splicing mechanism. These findings may provide valuable tools for a further analysis of *dmyc cis*-elements and mechanism of *dmyc* tight regulation at transcriptional level.

**Zusammenfassung**

Eine der wichtigsten Fragen in der Biologie des streng regulierten *myc* Gens ist, wie sowohl seine hohe Expression in wachsenden und sich teilenden Zellen als auch seine verminderte Aktivität während der Differenzierung und in adultem Leben erreicht wird.

Das Produkt des Proto-Onkogens *myc*, ein nukleäres Phosphoprotein, spielt eine essenzielle Rolle in der Wachstumskontrolle eines jeden Organismus. Das Gen ist evolutionär konserviert, und die wichtigsten Strukturdomänen des Myc Proteins bestehen aus einer sequenzspezifischen DNA-bindenden Domäne, der sogenannten "basic region/helix-loop-helix/leucine zipper" (bHLHLZ) und einer N-terminalen Transaktivierungsdomäne. Diese beiden Motive sind essenziel für die Funktion von *myc* in der Zellwachstumskontrolle, der Regulierung der Apoptose, und der Zelldifferenzierung.

Den genauen Kontrollmechanismus des *c-myc* Gens in Vertebraten als ein "immediate early gene" zu verstehen, ist immer noch eine Herausforderung. Da *Drosophila* nur eine Kopie des *myc* Gens enthält, anstatt mindestens drei Isoformen wie in der Maus, und die Fruchtfliege ein gut etabliertes genetisches Modellsystem darstellt, ist es möglich die *cis*-regulatorischen Elemente in *dmyc* zu studieren.

Das Ziel dieses Projektes bestand aus folgenden Punkten: (i) Identifizierung der *cis*-regulatorischen Elemente in der 5'-UTR, der intragenischen Region, und in der 3'-UTR, die für die endogene Expression des *dmyc* Gens verantwortlich sind; (ii) Aufklärung des Mechnismus der *dmyc* Transkriptionstermination; (iii) Charakterisierung der möglichen Transkriptionsinitiationsseiten; (iv) Erläuterung des Mechanismus der *dmyc* RNA-Spleissung. Um diese Fragen zu beantworten wurde wie folgt vorgegangen: 1) Bioinformatik-basierende Analyse der nichtkodierenden Regionen des *dmyc* Gens, um die regulatorischen Elemente zu identifizieren und daraus die Exprerimente zu planen; 2) Studien der Aktivität des *lacZ*-Reporters unter der Kontrolle der regulatorischen Teile des *dmyc* Gens; und 3) Anwendung des 5' RACE zur Aufklärung des Spleissmechanismus. Daraus wurden die unten aufgeführten Resultate erzielt, die zu folgenden Schlussfolgerungen führten:

Durch den Vergleich eines 40-kb Referenz-DNA Fragmentes des *dmyc* Lokus auf dem X Chromosom, auf dem das *dmyc* Gen liegt und der orthologen Sequenzen von

12 sequenzierten Drosophiliden mittels dem phylogenetischen footprinting Bioinformatik Programm EvoPrinter und *cis*-Decoder wurden mehrere mutmassliche Enhancer-Regionen mit konservierten Sequenzen gefunden. Darunter befanden sich mehrere konservierte E-Boxen, welche bevorzugte Myc-Bindungsstellen des CACGTG und CACTTG Typs sind. In der Intron 2 Region, identifizierten wir neben zwei konservierten E-Boxen mehrere Cluster konservierter Sequenzenblöcke upstream des vorhergesagten intronischen Promotors. Zwei der identifizierten Cluster konservierter Sequenzen umfassen: (i) ein Wiederholungssequenz-Element (ATGTTGCCA) wobei der Kern (TGTTGC) sich dreimal wiederholt und (ii) ein Dead-Ringer für den HLHm-3-2-Enhancer (CGCGTGGGAAAA), in welchem sich die Konsensus-Bindungsstelle (GTGGGAA) für den *suppressor of Hairy wing su(Hw)* befindet. Im großen 3'-UTR-Bereich stellten wir Cluster konservierter Sequenzblöcke fest, aber keine E-Boxen. Eine Suche mit PROMOTER 2.0 (ein Bioinformatik Programm basierend auf neuronalen Netzwerken und genetischem Algorithmus), ergab eine potentielle Promotorregion (bezeichnet als P0) im äußersten Bereich upstream. Eine Suche mit DNASTAR Lasergene 9.1 (Modul: GeneQuest) identifizierte den P1-Promotor in den proximalen 5'-UTR-Sequenzen, das P2 Downstream-Promotorelement (DPE) in Intron 2 und drei potentielle Polyadenylierungssignale in der 3'-UTR.

Auf Grundlage der bioinformatischen Analysen und dem Auftreten geeigneter Restriktionsstellen in den nicht kodierenden Regionen wurden verschiedene Deletionskonstrukte hergestellt. Die Transgene wurden entweder in $y^1$ $w^{1118}$ oder attB Fliegenembryonen injiziert, um unabhängige transgene Fliegenlinien zu etablieren. Die Aktivität des *lacZ* Reporters wurde im dritten instar Larvenstadium im Gehirn, den Imaginalscheiben, Embryonen und Ovarien der etablierten Fliegenlinien analysiert.

Im Gehirn ist *lacZ* in einer Reihe von Zellen meist proximaler Teile beider Hemisphären und entlang des ventralen Ganglions exprimiert. Diese Zellen könnten neuronale Vorläuferzellen sein, die eine hohe Wachstumsrate und Zellteilung aufweisen bevor sie differenzieren. Die *lacZ*-Aktivität in den Zellen in den vorderen Teilen der Hemisphären könnte eine bestimmte Klasse von kontinuierlich proliferierenden Neuroblasten, sogenannte Mushroom Body Neuroblasten (MBNbs), sein. In der Flügelscheibe ist *lacZ* meist um den Flügel Beutel und in dem Notum Region exprimiert. In der Augenscheibe wird *lacZ* überwiegend anterior und posterior

der morphogenetischen Furche (MF), und in der Antennenscheibe in der Mitte und ringförmig um die Mitte exprimiert. In der Beinscheibe wird die Reporteraktivität in konzentrischen Ringen in der Mitte der Scheibe mit einem Loch in der Mitte beobachtet. Die einzelnen Zellen in den Bein- und Flügelscheiben mit lacZ-Färbung, könnten Vorläuferzellen für Sinnesorgane (SOP) sein, die sich vor der terminalen Mitose und Differenzierung aktiv teilen. Im frühen Embryo sind mütterliche Transkripte nachweisbar. Später während der Erweiterung der Keimanlage ist *dmyc* intensiviert im Mesoderm, mid-gut, Rachen und dem anal-Pad detektierbar. In den Ovarien wird die Aktivität von *dmyc* überwiegend in den Nährzellen, Eizellen und schwach an der Spitze des Germariums detektiert. Die beobachteten Muster in den getesteten Geweben korrelieren mit der endogenen *dmyc* Lokalisierung.

Die Expression mit dem Konstrukt, das nur die Sequenz des Intron 2 enthält, zeigt die Existenz einer downstream liegenden Initiationsstelle, welche die Transkription des Reportergens antreibt. Dies entspricht der Erkenntnis, dass viele der entwicklungsmäßig aktiven Gene modularen *cis*-regulatorischer Elemente enthalten, um Gene während verschiedener Entwicklungsstadien der Entwicklung ein- und auszuschalten.

Aus den erhaltenen Ergebnissen schließe ich, dass die meisten, wenn nicht alle der regulatorischen Elemente, die für die korrekte Expression von *lacZ* im larvalen Gehirn und Scheiben, Embryonen und erwachsenen weiblichen Gewebe erforderlich sind, innerhalb des getesteten 40-kb-Fragment des *dmyc* Lokus vorhanden sind. Weiter schließe ich, dass die Sequenzen einschließlich des Intron 1, der 5'-UTR, sowie etwa 100 bp upstream der Transkriptionsstartstelle des P1-Promotors ausreichend sind, um ein *dmyc*-ähnliches Muster der Reporteraktivität in ovariellen Oocyten, Nährzellen und im Embryo, jedoch nicht in larvalen Geweben zu erhalten. Schliesslich wird *dmyc* wie viele entwicklungsregulierte Gene wahrscheinlich von verschiedenen Transkriptionsinitiation-Einheiten transkribiert, welche eine sich weit upstream befindende regulatorische Region, ein TATA-Box enthaltender proximaler Komplex und ein TATA-loses downstream Promotor-Element zusammen mit einem Initiator innerhalb der Intron 2 Region enthält.

Analyse der Polyadenylierungssignale, Poly (A)1, Poly (A)2, und Poly (A)3, am *dmyc* 3'-Ende ergaben, dass das *dmyc* Gen unterschiedliche Transkripte mit kürzeren und längeren Längen erzeugen würde, was den Angaben aus der Literatur entspricht.

Die 5' RACE-Analyse der cDNA-Enden von *dmyc* legt die Transkriptionsstartstelle am P1-Promotor 18 Basenpaare upstream des Starts der bekannten EST GM01143 innerhalb des 5'-UTR fest, wobei ein A als die erste Base transkribiert wird. Die Daten zeigen, dass die erste TATA-Box, welche zuvor rechnerisch vorhergesagt wurde, genutzt wird, um *dmyc* mRNA voller Länge zu generieren. Das größte Transkript enthält alle drei Exons, das nach der Entfernung der Introns durch konstitutiv regulierten Spleißmechanismus erzeugt wird. Diese Erkenntnisse können sich für eine weitere Analyse der *cis*-Elemente von *dmyc* und dessen strenger Regulierung auf Transkriptionsebene als nützliche Werkzeuge erweisen.

## 1. Introduction

For a fertilized egg to grow into a fully developed organism, with all its tissues and organs, high rates of cell growth and proliferation, apoptosis, cell adhesion, cell fate specification, and differentiation is required. All these processes need the activity of genes whose products are involved in regulating cell metabolism, biogenesis, and over all energy household of every single cell. All crucial aspects of early development are positively regulated by *myc* gene. Myc's action integrates extracellular signals into gene-regulatory programs within the cell to coordinate controlled growth (Grandori et al., 2000; King et al., 1986; Levens, 2003). The evidence for the strong involvement of Myc in early development comes partly from the intensive research work on stem cell biology. Intensive research of reprogramming somatic mammalian cells into induced pluripotent stem cells (iPSCs) resulted in the Nobel Prize for Physiology and Medicine in 2012 (Sir John B. Gurdon and Shinya Yamanaka). The generation of iPSCs was achieved by activation of four factors, including Myc as an essential factor (Figure 1) (Welstead et al., 2008). Furthermore, the importance of the crucial role of the *myc* during normal growth is evidenced by existence of 17,000-33,000 c-Myc binding sites in the human genome (Bieda et al., 2006; Cawley et al., 2004). Finally, it has been documented that c-Myc binds to 10-15% of human and fly genes (Orian et al., 2003; Zeller et al., 2006), indicating its high involvement in gene regulation as a transcription factor.

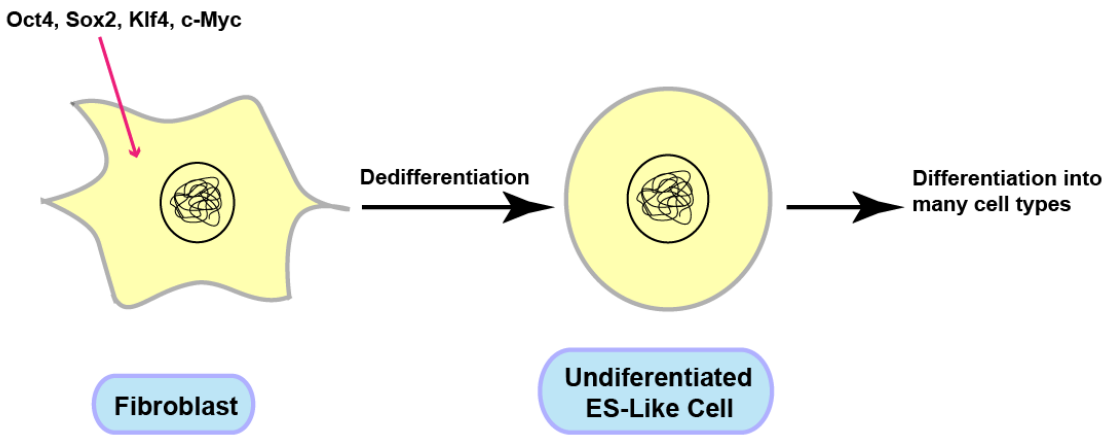



**Figure 1. Dedifferentiation of a differentiated cell.** Addition of four specific regulatory factors (Oct4, Sox2, Klf4, and c-Myc) is sufficient to reprogram a differentiated fibroblast towards pluripotency. The embryonic stem cell-like cell can potentially take the fate towards developing many cell types.

## 1.1 Developmental Patterning of Myc and its Function

The biological activity of the *myc* gene, as a gene regulatory protein and growth inducer, is required for blocking cell differentiation, induction of apoptosis, and progression through cell cycle (Amati, 2001; Cole and Henriksson, 2006; Patel and McMahon, 2006). Members of *myc* family gene play a central role during embryonic and post-embryonic development. For example, experiments with *Xenopus laevis* eggs have shown that *c-myc* m-RNA patterning is maintained in the egg, is detectable throughout the early cleavage stages of embryogenesis, and is accumulated in neural, muscular and mesodermal primordial tissues (King et al., 1986). In terminally differentiated cells Myc protein is nearly absent; in adults *myc* expression is confined to cell proliferation in the dividing cells of the tissues and during regenerative processes (Chin et al., 1995; Sodir and Evan, 2009; Zimmerman et al., 1986).

As an evolutionarily conserved basic region/helix-loop-helix/leucine zipper (bHLHLZ) transcription factor, Myc is a master regulator of cell growth and proliferation (Evan and Littlewood, 1993; Guo et al., 2009). Upon dimerization with Max, another bHLH protein, Myc binds to the E-box sequences of target genes to activate cellular growth and cell cycle progression (Figure. 1.1) (Blackwood and Eisenman, 1991).

Conversely, heterodimers of Myc and other Myc-associated zinc finger proteins, such as Miz1, can act negatively to regulate transcription of genes responsible for cell cycle arrest (Adhikary and Eilers, 2005; Eilers and Eisenman, 2008; Herkert and Eilers, 2011). Myc proteins can link growth with cell cycle progression via activation of the S phase cyclins, which are required for DNA replication (Dominguez-Sola et al., 2007). Regulation of cell growth and division is critical for animal development because too little growth leads to small organs and small body size, whilst excessive growth can lead to tissue overgrowth, initiation of cancer, or perturb normal growth (Adhikary and Eilers, 2005; Dang, 1999; Eisenman, 2001; Schmidt, 1999).

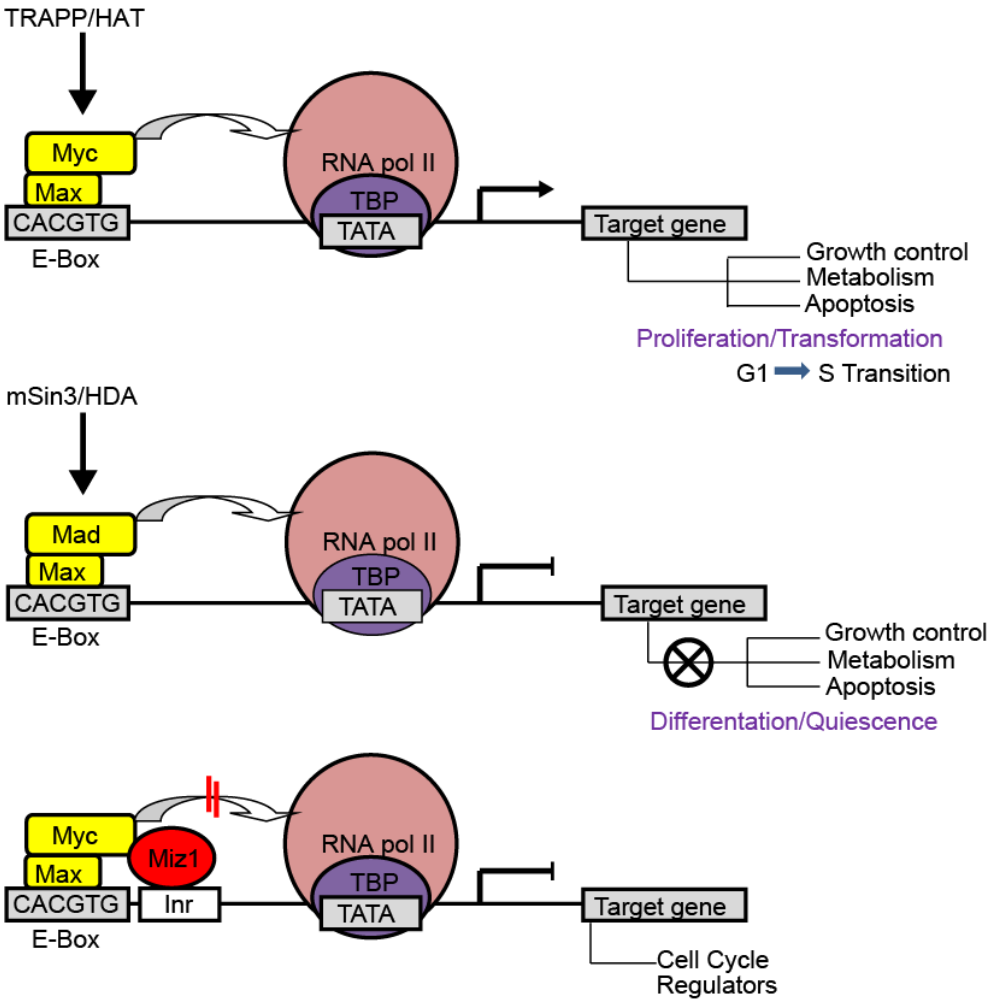


**Figure 1.1.** Simplified schematic interaction of Myc/Myc-associated bHLH proteins at the promoters of target genes. Upon binding to E-box sequences in the promoter region of target genes, heterodimers of Myc-Max can recruit chromatin remodeling complex TRAPP/histone acetyl transferase and interact with the bound basal transcription machinery at the TATA region of target genes to activate transcription. Conversely, heterodimers of Mad/Max transcription factors recruit mSin3/histone deacetylases to counteract Myc activity and repress Myc target genes by regulating differentiation and cell cycle arrest. The binding of Myc/Max dimers can interfere with the function of transcription activator Miz-1 to inhibit the recruitment of cofactor proteins like p300 to the promoters of genes responsible for cell cycle regulation (Kharazmi et al., 2011).

### 1.1.2 Myc's Oncogenic Role in Cancer Development

As an immediate early and cancer-critical proto-oncogene *myc* plays a crucial role in malignant transformation. In the history of neoplastic diseases, it was found that the overexpression of *v-myc* was the cause for cancer development by avian retroviruses MC29 and AMV (Kelloff and Vogt, 1966). Thereafter, the identification of *myc* in tumor tissues was the first example of insertional mutagenesis in cancers such as B-cell lymphomagenesis in chickens (Marcu et al., 1992). Further intense studies have located the mammalian *c-myc* locus at the breakpoints of chromosome translocations with immunoglobulin loci, in Burkitt's lymphoma in humans and plasmacytomas in mice (Marcu, 1987; Marcu et al., 1992).

Since then, intensive analysis on oncogenic characters and developmental aspects of *myc* have identified further members of the *myc* family proteins, namely N-, L-, and S-Myc (Benassayag et al., 2005; Cotterman et al., 2008; Nau et al., 1985). In terms of myc's tumorigenic function, it has been shown that *myc* possesses the ability of oncogenic collaboration with other oncogenes including *ras*, *bcl-2*, *LPA*, *Bcl-xL*, *Mcl-1*, and *cyclin D1* (Bissonnette et al., 1992; Fanidi et al., 1992; Sears et al., 1999; Taghavi et al., 2008).

Chromosomal translocations and genetic alterations result in the deregulation of tight control of *myc* at the transcriptional and posttranscriptional level. Alterations in the regulation of *myc* family members are observed in hematopoietic and nonhematopoietic tumors (Leder et al., 1983; Popescu and Zimonjic, 2002).

### 1.1.3 Transcription Factor Myc and Signal Transduction Pathways

The phosphoprotein Myc belongs to a small family of highly related bHLHLZ transcription factors. The major part of the polypeptide open reading frame is located in the second and third exons. The first exon of the *myc* genes is not conserved among *myc* family members, but it contains *cis*-regulatory elements. Among these, the P1 and P2 promoters have been definitively identified in the murine and human *c-myc* (Wierstra and Alves, 2008). Human, murine and avian c-Myc contain the two major polypeptides, Myc1 (454aa) and Myc2 (439aa), and one less predominant form MycS (339aa) that appear to be translated from multiple initiation sites (Benassayag et al., 2005; Luscher and Eisenman, 1990). c-Myc isoforms are functionally distinct with Myc1 abundant during early growth phase, Myc2 predominant when cells reach growth arrest. However, c-MycS accumulation is not persistent and it seems to be capable of both, negatively regulating the activity of Myc1 and Myc2, but also retaining proliferation potential (Benassayag et al., 2005; Spotts et al., 1997).

Main functional domains of c-Myc1 reside within the N-terminal domain (NTD) and C-terminal domain (CTD) (Figure 1.1.3). Major transcriptionally functional and evolutionarily conserved domains in NTD include Myc-homology Box 1 (MB1) and Box 2 (MB2). MB1 contains major phosphorylation sites that regulate transcription and transformation, and it is a hotspot for mutation in cancer (Henriksson and Luscher, 1996; Sakamuro and Prendergast, 1999). The C-terminal domain contains a BHLHLZ region that mediates DNA binding and protein-protein interaction (Figure 1.1.3). In addition to heterodimerization with Max, another BHLHLZ protein, it has been shown that many factors such as Nmi, YY-1, AP-2, TFII-1, CRCA1, and Miz-1 interact with this region (Sakamuro and Prendergast, 1999).

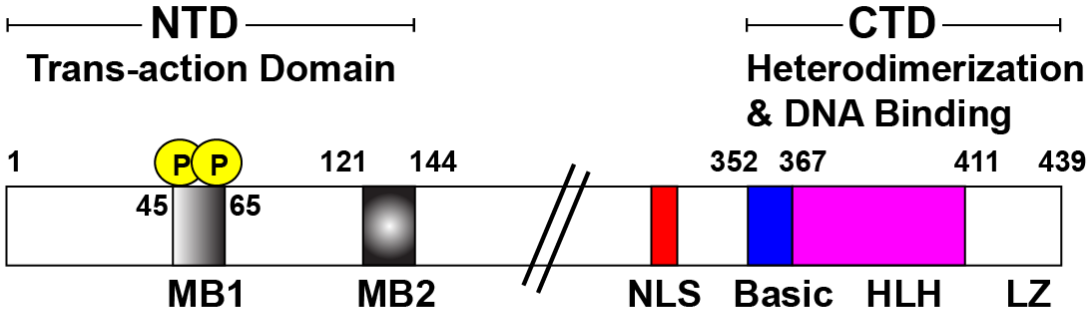



**Figure 1.1.3. The structure of the human c-Myc protein.** The N-terminal domain contains highly conserved Myc-homology Box1 (MB1) and Box2 (MB2), important for its transcriptional activity. Functionally critical phosphorylation sites at S62 and T58 reside in the MB1. The C-terminus (CTD) includes nuclear localization signal (NLS) and DNA binding domain. HLH: helix-loop-helix, LZ: leucine zipper. (After Sakamuro and Prendergast, 1999).

Myc is at the junction of virtually all biological functions involved in cell homoeostasis. Myc influences multiple pathways ranging from proliferation, apoptosis, and differentiation to biogenesis. In response to mitogenic stimuli Myc induces DNA replication via activation of the S Phase cyclins to promote cell growth and progression through cell cycle. Many signaling pathways respond to Myc activity. Smad proteins belong to the transforming growth factor beta superfamily (TGF-β) signaling that keeps control of aberrant proliferation and commences cell differentiation. It has been proposed that TGF-β represses transcription of *c-myc* by direct binding of Smad3 to a novel repressive Smad binding element in *c-myc* promoter region (Anderson et al., 1981; Breathnach and Chambon, 1981). Additionally, it has been shown that *c-myc* expression is reduced in normal epithelial cells in response to the growth inhibitory cytokine TGF-β (Kollmar and Farnham, 1993; Lo and Smale, 1996).

A large body of data indicates that *c-myc*, as a developmentally regulated gene, is direct downstream target of Notch family proteins. Notch pathway is a major regulator of cell fate specification and pattern formation during development (Maillard et al., 2005). Notch signaling triggers its action in many different ways that is dependent on cell type, type of Notch factor, and dose. Mutation in the oncogene Notch1 induces constitutive expression of *c-myc* and it contributes to the growth of T-cell acute lymphatic leukemia (T-ALL) (Aster, 2005; Weng et al., 2006).

The G-protein-coupled receptor Ras/Raf-signaling pathway has been documented to activate the expression of *myc*. It has been shown that upon binding of a mitogen to a G-protein receptor and activation of Ras/Raf cascade in serum starved fibroblasts the expression of the *c-myc* gene is induced within 2-6 h (Kerkhoff et al., 1998).

It has been implicated that the transcription factor T-cell specific factor (TCF), a component of Wnt signaling, upregulates the activity of *myc* gene (He et al., 1998; He et al., 2009). Myc is in turn required for the majority of Wnt target genes activation. Furthermore, a study in adult mice has shown that loss of adenomatous polyposis coli tumor suppressor protein (APC) with simultaneous deletion of *myc* rescues development of murine intestinal cancer (Baldwin et al., 2003).

Myc is involved in promoting apoptosis at multiple junctions by acting on death receptor pathways. Upon deprival of survival factors, *c-myc* expression induces apoptosis that is mediated by tumor-suppressor gene product p53 (Hermeking and Eick, 1994). Another study indicates that cooperation between p53 and Myc to induce apoptosis is dependent on the induction of Bax to enhance p53-mediated apoptosis. However, this relationship between Myc and p53 is necessary but insufficient to trigger apoptosis (Taylor et al., 2006). Further factors have been discovered in cooperating with the apoptotic roles of Myc. Among these, connections to CD95/Fas and TNF pathways, and involvement of tumor suppressor protein p19ARF, and the c-Myc interacting adaptor protein Bin1 in inducing apoptosis have been implicated (Prendergast, 1999).

These are just a few examples of the large list of the signaling pathways in which the involvement of Myc is vital for cell normal growth and function. The crucial role of Myc in cell behavior reinforces the importance of Myc's tight control at all levels.

## 1.2 Control of the *c-myc* Promoter

The regulation of *myc* gene is very complex and the investigation of its physiological function reveals that *myc* is involved in many cellular processes. Since its molecular cloning and the subsequent identification of its cellular homologues, *myc* has been intensely studied, with > 25'000 publications on it currently found in the biomedical research literature. Nevertheless, the exact mechanism controlling expression of the mammalian *c-myc* gene is still enigmatic and a challenge (Wierstra and Alves, 2008).

Four promoters, P1, P2, P3, and P0 have so far been identified in the human *c-myc* gene (Figure 1.2) (Battey et al., 1983; Marcu et al., 1992; Ray and Robert-Lezenes, 1989; Spencer and Groudine, 1991; Watt et al., 1983). P1 and P2 both contain TATA box; P0 and P3 are TATA less. The majority of transcripts initiate from P1 and P2, with a clear dominance of P2, giving rise to 75-90% of cytoplasmic *c-myc* mRNA. Transcription initiation from P1 produces 10-25%; less than 5% of the transcripts stem from P0 site, and also less than 5% from P3 (Bareket-Samish et al., 2000; Yean and Gralla, 1997).

Numerous gene regulatory protein binding sites have been identified in the noncoding regions of the *myc* gene. Performance of DNase I hypersensitive (DH) assays have confirmed many protected sites throughout the *c-myc* gene body as well as in the upstream and downstream flanking sequences (Ishihara et al., 2006; Mautner et al., 1995; Murphy et al., 1996) (Figure 1.2). The structure of euchromatin, transcriptionally active chromatin, differs from the inactive and heterochromatic regions of the DNA. The active DNA, which includes the coding portions as well as upstream and downstream adjacent nontranscribed regions of the DNA are sensitive to nuclease digestion. The DH sites in the *c-myc* locus correlate with the presence of regulatory units and binding sites that are specific to DNA-binding proteins (Murphy et al., 1996).

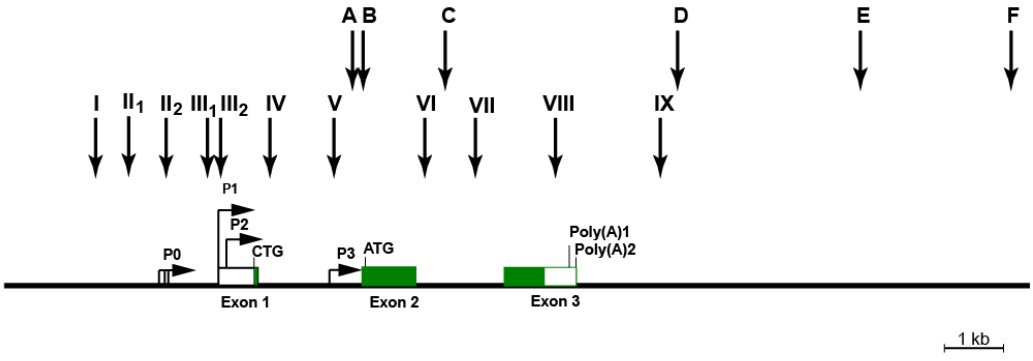



**Figure 1.2. Structure of the human *c-myc* gene with the depiction of DNase hypersensitive sites.** Human *c-myc* gene includes three exons, (white and green boxes), four promoters (P0, P1, P2, P3), two major translation start sites (CTG and ATG), two polyadenylation signals [poly(A)1, poly(A)2]. Black arrows show DNase I hypersensitive sites. For DH sites Roman numerals (I-V) see Bentley and Groudine, 1986a; for the sites (V-IX) see Murphy, et al., 1996; and DH sites (A-F) see Mautner, et al., 1995. (DH sites A, B, C are named VI, VII, VIII, and DH sites D, E, F are named 1.5, 6.5, 9 in Mautner, et al., 1995).

P0 contains multiple initiation sites. The murine and rat *c-myc* genes contain the promoters P1, P2, and P3, but lack P0.

## 1.3 *dmyc*, *Drosophila melanogaster* Homolog of Human Proto-Oncogene, *c-myc*

The complex and tight regulation of *c-myc* in different biological context reflects its important and enigmatic role in normal tissue homoeostasis. Since its identification, many principles of its regulation and biology have been answered; yet, many outstanding and challenging questions remain to be resolved. Except vertebrates and species down to *Drosophila* in the animal kingdom no other species has been reported to contain a copy of *myc* gene.

*Drosophila myc* was identified by using human Max to screen a two-hybrid library prepared from *Drosophila* cDNAs (Gallant et al., 1996). Many functionally important regions of c-Myc are conserved in dMyc (Figure 1.3): (i) the NH2-terminal region with 57% identity to the human c-Myc box 2, which contains the conserved sequence DCMW; mutations in DCMW abrogate Myc activity (Brough et al., 1995); (ii) the centrally located acidic region with 57% identity to the “acidic region” of vertebrate c-Myc; (iii) the COOH-terminal segment with 40% identity to the human c-Myc BHLHLZ domain (Figure 1.3).

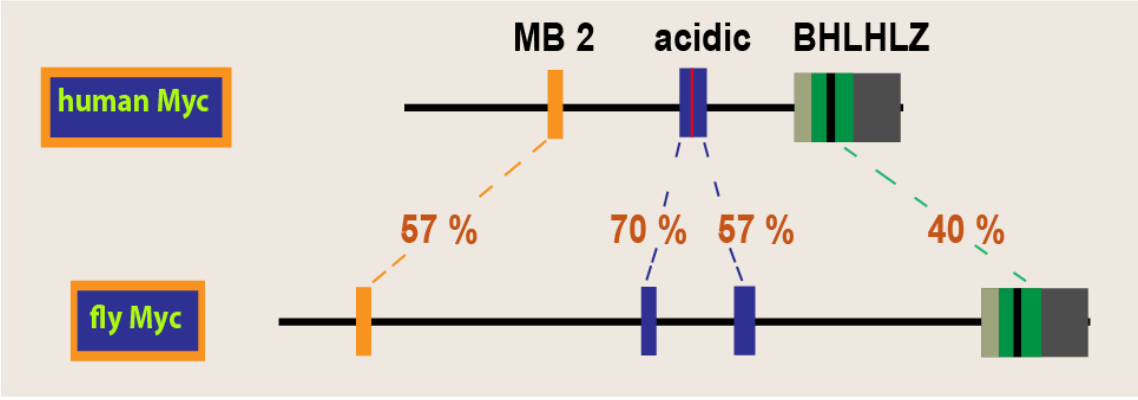



**Figure 1.3. Evolutionarily conserved functional regions between human c-Myc and *Drosophila* dMyc and percent similarity.** As illustrated, the Myc box 2 (MB 2), the acidic region, and the basic-helix-loop-helix region (bHLHLZ) of c-Myc are conserved in dMyc. Intervening regions show less similarity (picture adapted from my diploma work).

A variety of biological activities have reinforced functional conservation between *Drosophila* dMyc and human c-Myc, such as the ability of dMyc to drive the cell cycle in *c-myc* null fibroblasts and to transform primary mammalian cells (Schreiber-Agus et al., 1997; Trumpp et al., 2001). Conversely, c-Myc can rescue lethal mutations of *dmyc* (Benassayag et al., 2005). The evolutionarily conserved structure and function between dMyc and c-Myc has prompted the use of *Drosophila* as a model to gain insight into many aspects of mammalian Myc biology. In particular, *Drosophila* genetic models have demonstrated that dMyc controls cell growth and cell division by regulating its direct targets involved in protein biogenesis and metabolism, as does c-Myc (de la Cova and Johnston, 2006; Trumpp et al., 2001). Both, *Drosophila* Myc and c-Myc heterodimerize with their partner Max protein and recognize the same binding sequence, CACGTG, to activate transcription.

## 1.4 *Drosophila melanogaster* Larval Primordial Tissues: Relevance to this Study

In 1909 Thomas Hunt Morgan was the first geneticist who introduced the fruit fly *Drosophila melanogaster* in the field of genetic studies. Since then the short life cycle of *Drosophila*, its profoundly established genetic tools, and transparency of its tissue primordial makes the model organism *Drosophila* a favorite and convenient system to follow many developmental patterning processes, gene expression and disease development.

After midembryogenesis, and by early stage 17, some of the larval epidermal primordial invaginate as imaginal discs that are connected to larval epidermis only by a thin peripodial stalk (Cohen et al., 1991; Simcox et al., 1991). These primordial tissues give rise to many different tissues in the adult fly as the embryo undergoes metamorphosis (Figure 1.4) (Jeibmann and Paulus, 2009). The imaginal discs and larval brain are important primordial tissues that involve cell proliferation, apoptosis, differentiation and growth in order to undergo the process of organogenesis. An accumulating body of evidence supports the involvement of *myc* in the above processes (Amati and Land, 1994; Arabi et al., 2005; Blackwood et al., 1992; Brough et al., 1995; Grewal et al., 2005; Levens, 2003; Marcu et al., 1992; Mautner et al., 1996; Orian et al., 2003; Oster et al., 2002; Patel and McMahon, 2006; Sakamuro and Prendergast, 1999; Weinberg, 1995), making the imaginal discs and the brain

interesting tissues to examine the expression pattern of reporter gene under *myc* control.

Different stages of embryonic development facilitate their study because they are transparent, possess short metamorphosis time spans, and are easy to manipulate. Already in the very early stages of development, the expression and repression of each developmentally regulated gene is restricted to different anteroposterior and dorsaventral compartments of the embryo, which leads to fate map specification of the cell and formation of the three germ layers: endoderm, mesoderm, and ectoderm (Bate, 1990; Rivera-Pomar and Jackle, 1996).

Ovarian tissue is primarily a derivative of follicular epithelium cells that undergo dynamic cycles of cell proliferation (Dobens and Raftery, 2000). The egg chamber, surrounded by follicle cells, include nurse cells and the oocyte (Jaglarz et al., 2008). High proliferation rate, diverse cell types, and migratory processes that require the involvement of many master regulators, make the ovary a favorite system to study biology of *myc* gene as “immediate early gene”.

To sum up, since *myc* gene is dynamically patterned in all the above mentioned tissues during early development, where cells actively cycle, in this study I looked for the activity of reporter *lacZ* under control of different regulatory regions of *dmyc* in third instar larval brain, different imaginal tissues, during embryogenesis, and in adult female ovarian tissues.

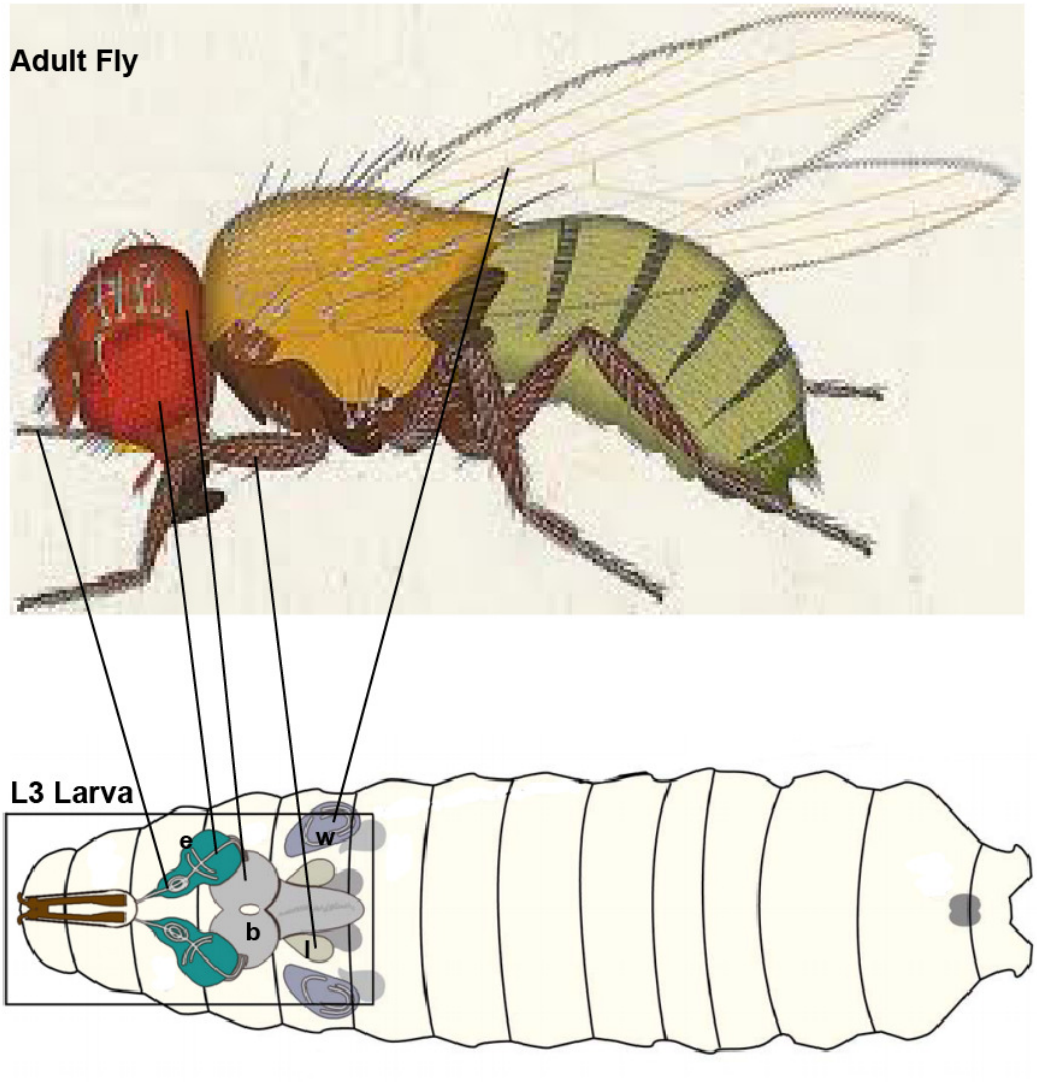



**Figure 1.4. Third instar larva with imaginal discs corresponding to tissues in the adult fly. e),** eye-antennal disc; **b),** brain; **w),** wing disc; **l),** leg disc. (Adapted from Hartenstein, V. Atlas of *Drosophila* Development; Jeibmann and Paulus, 2009).

## 1.5 Developmental Expression Pattern of *dmyc*

*dmyc* is synthesized in a dynamic spatial and temporal pattern during development of *Drosophila*, as determined by *in situ* hybridization and Northern blot analysis (Gallant et al., 1996; Johnston et al., 1999). In third instar larval brain *dmyc* activity is ubiquitous, with an increased level of expression in the distal and middle parts of the lobes and within dividing neuroblasts in the middle parts of the ventral ganglion (see Results, section 2.5.1). Additionally, *dmyc* expression is confined to a limited number of cells distributed in the two proximal halves of the hemispheres and along the

ventral ganglion (Kharazmi et al., 2011). RNA *in situ* hybridizations have shown *dmyc* mRNA from third instar larvae to localize in the wing disc around the wing pouch and to the notum (Johnston et al., 1999). In the eye disc, *dmyc* is synthesized posterior and anterior to the morphogenetic furrow (MF), but not in the MF (L. Quinn, personal communication). In antennal discs *dmyc* expression is mainly detectable around the center (see Results, section 2.5.1 and Figure 2.5.1.a). In the leg disc endogenous *myc* is expressed around the middle of the disc with the center lacking *dmyc* activity.

Unlike *dmax*, which is expressed at constant level throughout development, *dmyc* expression is detectable from early stages of oogenesis; in the oocyte maternal transcripts are detectable, and later the zygotically derived transcripts can be detected in various stages of embryogenesis. At later stages *dmyc* expression is intensified in pharynx, mesoderm, and during germband extension in the midgut and anal pad. The strong staining is persistent until midembryogenesis (Gallant et al., 1996; Kharazmi et al., 2011) . In adult females, *dmyc* mRNA is highly expressed in the oocyte and nurse cells, but is weakly expressed at the tip of germarium (see Results, section 2.5.1 and Figure 2.5.1.b).

### 1.6 Approaches Used in the Study for the Investigation of the *dmyc* Promoter

Two main tools in molecular biology namely, reporter activity study and bioinformatics analysis can be used to localize the *cis*-regulatory regions of a gene (Kipp and Mayo, 2009; Yavatkar et al., 2008). Intensive studies and computational approaches have been undertaken in dissecting human *c-myc* promoter, as comprehensively summarized in Wierstra (2008). The studies have resulted in identification of a variety of repressor and activator binding sites in the enhancer regions throughout the gene body. The investigations have further shown that the *c-myc* gene regulatory region is modularly structured; it contains four promoter (P1, P2, P3, and P0) and two polyadenylation sites (Marcu et al., 1992; Spencer and Groudine, 1991). However, it has been very difficult to characterize enhancer elements responsible for correct expression of *c-myc* gene. In one of the *in vivo* studies Mautner et al. (1996) tested constructs including up to 50-kb of the *c-myc* locus, but the constructs failed to express exogenous *c-myc* gene substantially. The insertion of the immunoglobulin κ-intron and 3' enhancers, however, activates *c-myc* transcription when placed adjacent to or separated from the *c-myc* promoters by as much as 30-kb (Mautner et

al., 1996). The coupling of these enhancers with the *c-myc* gene mimics the chromosomal translocation in Burkitt's lymphoma, in which the *c-myc* gene gets translocated to the immunoglobulin loci. Although the distance of the chromosomal breakpoints from the *c-myc* locus varies from 1-kb to 350-kb, the translocated *c-myc* becomes actively transcribed from the P1 promoter.

The above findings indicate that the examined sequences of 50-kb contain functional promoter elements but lack at least some essential enhancer elements. Given the large size of 160 to 350-kb, the identification of additional enhancer elements in the *c-myc* region is expected to be difficult (Gombert et al., 2003; Mautner et al., 1996). Such studies can more efficiently and conveniently be performed by investigating *myc* regulation in the fruit fly for several reasons: (i) *Drosophila* possesses well-established genetic tools; (ii) *dmyc* is the sole homolog of *c-myc* in the fly; (iii) *dmyc* locus spans over 40 to 50-kb, instead of 160 to 350-kb, (iv) life cycle of *Drosophila* is much shorter than mouse life cycle.

### 1.6.1 Computational Approaches

The occurrence of conserved regions and repetitive sequence motifs in noncoding DNA has been of great value for the identification and characterization of *cis*-regulatory elements. The phylogenetic footprinting tools *EvoPrinter* (Yavatkar et al., 2008) and *cis*-Decoder (Brody et al., 2007) are efficient software tools that can readily serve for the identification of conserved sequence blocks in developmental genes. *EvoPrinter* facilitates the multialignment and rapid identification of evolutionarily conserved sequence blocks as they exist in the species of interest. The *cis*-Decoder then characterizes repeat motifs within the conserved sequence blocks and detects conserved elements among functionally related enhancers. It is important to mention that *EvoPrinter* and *cis*-Decoder do not detect polyadenylation signals or core promoter elements, such as TATA boxes.

The neural network genetic algorithm PROMOTER 2.0 can be used to predict promoter region in eukaryotic genes (Knudsen, 1999). PROMOTER 2.0, combined with the consideration of CCAAT or bHLH recognition motifs facilitates the recognition of discrete subpatterns that are recognized by RNA Pol II. The identified subpatterns can then be further analyzed by the bioinformatics tool DNASTAR laser Gene, module GeneQuest, to detect *cis*-acting transcriptional regulatory elements

such as GC box, 5'-GCGCGGC-3' (transcription factor SP1 binding site) (Song et al., 2009), TATA box (also called Goldberg-Hogness box) (Lifton et al., 1978), Initiator Element (transcription initiation site) (Smale et al., 1998; Smale and Kadonaga, 2003), Downstream Promoter Element (DPE) (Smale, 2001) and transcription termination signals in the 3'-UTR region.
The data obtained by computational methods serve as a supportive tool for the generation of overlapping deletion constructs covering the entire noncoding regions of a gene that can be used to perform reporter activity studies.

### 1.6.2 Reporter Activity Study

Reporter gene assay is an effective and easy method for the measurement of the activity of a particular promoter in vivo. A common reporter is *E. coli lacZ* gene, which resides within the *lac* operon and encodes the β-galactosidase (β-gal) enzyme. The enzyme is 120 kDa and for its activity forms a tetramer. In bacteria the enzyme cleaves lactose to glucose and galactose to provide the cell with the source for carbon and energy.

*In vitro* the synthetic compound 5-bromo-4-chloro-indolyl-β-D-galactopyranoside (X-gal) consists of galactose linked to an indole molecule; it can be cleaved by the enzyme β-gal. The insoluble cleavage product, an indoxyl glycoside has a blue color and is indicative for the presence of β-gal enzyme (Kiernan, 2007). The accurate detection of the enzyme β-gal in the system allows the use of *lacZ* reporter as a measurement for promoter activity.

Two major transgenesis approaches for the study of reporter activity in *Drosophila melanogaster* have been established, namely random P-element transformation (Rubin and Spradling, 1983) and site-directed phage ΦC31 integrase transgenesis system (Bischof et al., 2007; Venken et al., 2006). On the basis of computational analyses of the noncoding regions of a gene, both transgenesis systems can be used to introduce different truncations of the noncoding regions of a gene into the fly genome and study the expression of a reporter under control of a particular promoter. Random P-element transformation requires analysis of many independent lines to avoid insertion locus position effect on the promoter activity. Site directed mutagenesis takes advantage from the phage attachment sites attB and attP, and

allows the irreversible integration of a transgene into a particular landing site in the genome. This approach helps to save extra labor and avoids position variegation effects. Since the method is new in the field of reporter activity studies, a combination of both methods ensures more precise interpretation of the promoter activity.

## 1.7 Transcription Initiation from the *dmyc* 5'-UTR

Eukaryotic transcription involves a variety of different steps including pre-initiation, initiation, elongation, and termination. The first steps of transcribing a gene start with pre-initiation that is followed by stable initiation of transcription. Initiation requires the presence of a core promoter sequence in the DNA, namely the TATA box that is normally located about 50 bases upstream of the initiation site (Lee and Young, 1998) . An appropriate initiation can only be launched when the pre-initiation complex containing RNA polymerase II is stabilized at the TATA and activators/repressors, along with any associated cofactors, responsible for modulating transcription, are assembled (Dvir et al., 2001).

In the work here, using *Drosophila* adult poly A+ mRNA and amplification of the 5'-end of the *dmyc* cDNA by 5' RACE, I have identified the start of the *dmyc* largest transcript (Results, section 2.6).

## 1.8 Splicing Mechanism of the *dmyc* Large Transcript

In contrast to prokaryotes, the eukaryotic transcription and translation occur in two completely separate systems. RNA polymerase II transcribes protein coding genes; and its product, pre-mRNA, is much longer than the mature RNA. Pre-mRNA undergoes three maturing processes in parallel to transcription prior to its export from nucleus into cytoplasm for translation: (i) removal of intervening noncoding introns and splicing of exons; (ii) capping at its 5'-end; (iii).addition of a poly-A tail at its 3'-end (Green, 1986, 1991).

Intron removal and exon joining in maturing mRNAs is an extremely precise and energy consuming mechanism. All the intervening noncoding sequences must be removed at the level of nucleotide precision before the protein coding exons are joined together. RNA splicing is achieved in two main ways: it includes constitutive or

regulated splicing, in which the introns are removed and the exons on each side of the intron are spliced. Differential or alternative splicing occurs by joining of one splice site to one of several complementary splice sites so that variants of protein products from the same mRNA molecule are generated (Green, 1986).

One of the methods for the investigation of the splice events in a particular gene is examining the cDNA that has originated from the mature mRNA. Rapid amplification of full-length cDNA ends (RACE) that has been reverse transcribed from the mRNA, delivers PCR products that can be sequenced to determine the splice junctions in the cDNA (Krug and Berger, 1987). The obstacle of the strategy is to overcome the difficulty with the RNA instability since RNA is single stranded and is subject to rapid degradation. Drummond et al. (1985) showed in their study done with mRNA molecules from *Xenopus* oocytes that addition of a poly (A) tail to 3'-end of RNAs increases the stability at least for 50%. The injected poly A+ mRNA into oocyte was detectable in a considerable amount even after 48 hours of injection. In addition to poly (A)-tailing, it is well known that ligation of an oligonucleotide to the 5'-end of the mRNA before reverse transcription reaction, ensures generation of a full-length cDNA (Krug and Berger, 1987).

In the study here, using *Drosophila* adult poly A+ mRNA I amplified the *dmyc* full-length cDNA ends by 5' RACE, and have shown that the *dmyc* largest transcript is produced by regulated splicing (see, Results section 2.7).

# 2. Results

## 2.1 *dmyc* Gene in Twelve Sequenced *Drosophila* Species

Genome sequencing has yielded valuable data that can be applied to multi-genome comparative studies for purposes like estimation of evolutionary distance in the animal kingdom or identification of putative *cis*-regulatory regions important in gene regulation (Reneker et al., 2012; Yavatkar et al., 2008). In the study reported here, I was looking for *dmyc* regulation at the molecular level that is conserved in evolution. Therefore, I used the DNASTAR Lasergene 9.1 (module: MegAlign) to curate 40-kb genomic sequences of twelve *Drosophila* species for the *dmyc* genomic organization. For defining the genomic structure of the *dmyc* gene in 11 species, the *Drosophila melanogaster myc* gene, *dmyc* (FlyBase Annotation ID CG10798, also known as *dm*), could serve as reference DNA. The *dmyc* gene is about 12.8-kb in size, is located on the X chromosome, and is flanked by two other genes, the 5' *dunce* and the 3' *CG12535* gene (Figure. 2.1.a).

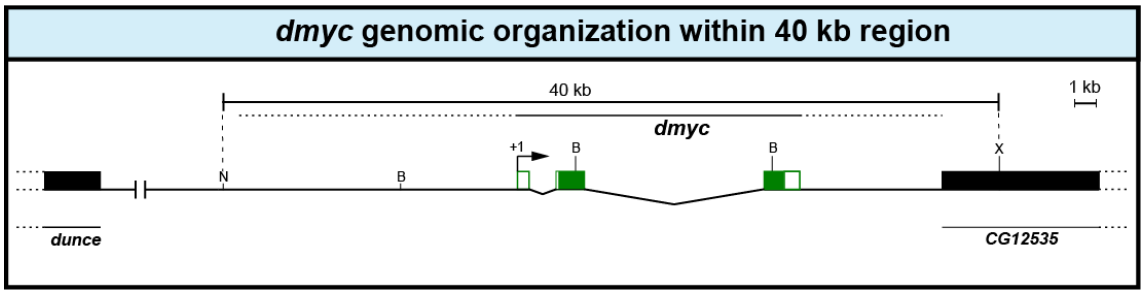



**Figure 2.1.a. The examined 40-kb *D. melanogaster dmyc* genomic locus is depicted.** The 40-kb genomic fragment on the chromosome X of *Drosophila melanogaster* resides between nucleotides -3253633 to +3293632, and harbors the *dmyc* locus (nucleotides -3267216 to +3280049). The 5' flanking gene *dunce*, also annotated as *CG32498*, is located about 33-kb upstream of the examined 40-kb sequences. The 3' flanking gene, *CG12535*, is located about 6.5-kb to the 3'-end of the *dmyc* gene (nucleotide +3280049).

Analysis of the *dmyc* gene structure shows, except for *D. pseudoobscura* that did not get rid of its long introns throughout the evolution, the size of the *dmyc* gene and its genomic organization is very similar among the 12 sequenced *Drosophila* species (Figure 2.1.b). In most of the twelve species the *dmyc* gene contains two introns (black line in the gene body) and three exons (green box).

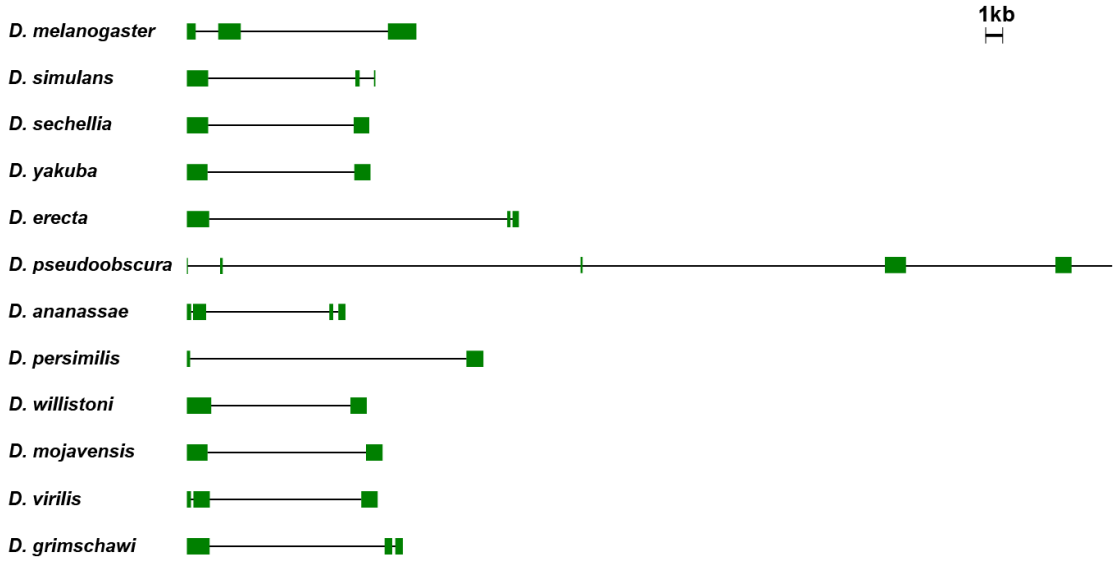



**Figure 2.1.b. Structure of the *dmyc* gene in twelve *Drosophila* sibling species.** Most of the *dmyc* genes in the sibling species show similar organization of the gene structure. Using computational comparative searches of the 40-kb region spanning the *dmyc* gene, multiple conserved sequence blocks were detected that are common to most species (see, Results section 2.1.1).

### 2.1.1 Computational Comparative Analyses of the *dmyc* Locus: Prediction of Conserved Sequence Blocks within the 40-kb *dmyc* Locus

It is known that noncoding DNA contains evolutionarily conserved sequence blocks that may refer to *cis*-regulatory elements. *EvoPrinter* (Yavatkar et al., 2008) is a comparative genomics tool that can be used to identify multispecies conserved sequences (MCSs) in orthologous DNA from sibling species. A good *EvoPrint* of the noncoding region can be processed by *cis*-Decoder (Brody et al., 2007) to identify *cis*-regulatory DNA of a particular species.

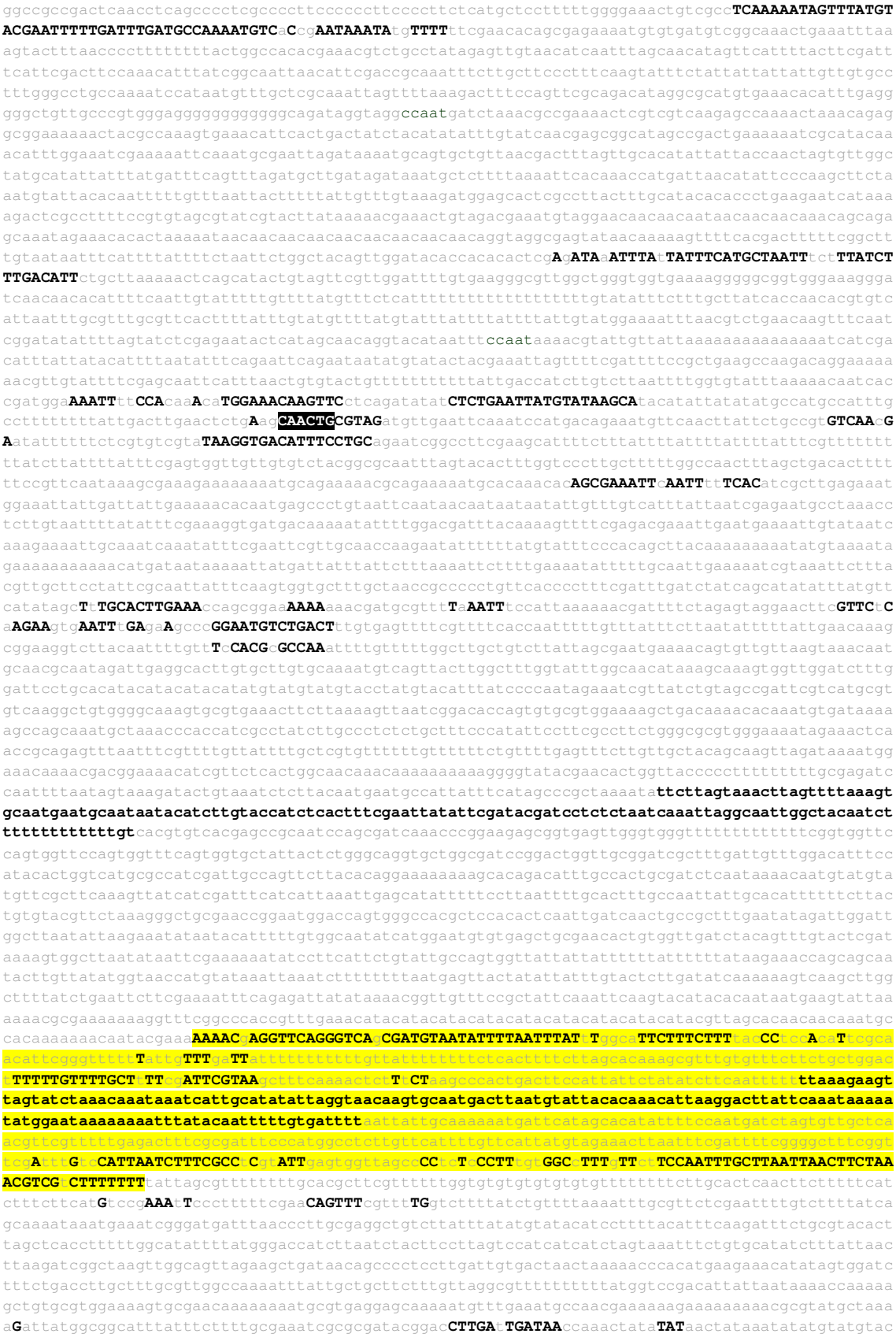
To identify *cis*-regulatory regions at the *dmyc* locus, a 40-kb fragment on the X chromosome harboring the *dmyc* gene of *Drosophila melanogaster* was used as the reference DNA to align orthologous sequences from 12 sequenced Drosophilids to test for the existence of multispecies-conserved sequence blocks in the noncoding regions. We identified several putative enhancer regions with conserved sequences, including several conserved E-boxes, which are preferred Myc binding sites (Blackwell et al., 1990; Jones et al., 1996) of both the CACGTG and CACTGT type

(Figure 2.1.1.a and Figure 2.1.1.b). The identified E-boxes may play an important role in the *dmyc* autoregulation events (see Discussion, section 3.1.1).

Comparison of the sequences from intron 2 identified multiple clusters of conserved sequence blocks upstream of the predicted intronic promoter (Figure 2.1.1.b). In addition to two conserved E-box sequences (CACGTG and CACTTG), two of the several identified clusters of conserved sequences include: (i) a repeat sequence element (ATGTTGCCA) where the core (TGTTGC) is repeated three times (Figure 2.1.1.b), (ii) a dead-ringer for the HLHm-3-2 enhancer (CGCGTGGGAAAA), within which the consensus binding site (GTGGGAA) for *suppressor of Hairy wing su(Hw)* resides (Figure 2.1.1.b).

In the large, approximately 10-kb 3'-UTR region, we identified clusters of conserved sequence blocks, but no E-boxes as was the case for the intronic region (Figure 2.1.1.b).

**C**hrX:3,253,853-3,261,899



atat**AT**actaaatttacctatagtgctgttatagcaaccaaacttaaaggccttcaaaacggcaaaaaacgaagacgaagaaaatattttagaa
aaacgaaaaagtgatagtaaattagtttagtttagtttgaattatactacgacacttaagactttgacaattgataagatttcattggggcaga
tatagatattgtgctgccaactaaatgtataacttatatatacgcacgtttctagatacggatagcatttttgctagatttcaagagaaaagg
ggaaacttaattagcaactgattaaaacggggaaatcacagcaaagcctcttacgactgcagttaactaatcatttgagatacaaaaaaaaaat
cctaattacgaaacggaatcaaaggatatttcaagtacaccatttcaaaaataatcaaattaaaatctaagcacacattttctattcccatcag
ctacataaatattgaactcatggtggaaaaaaaatcatcatccttggcgtgatttttaccctgacatttagtcgattgacttaaatatctccct
caacttttcttttcaaacgcaagcaaactcagcgagttgtatttagaaaagcgaccccagttggtctaaaattaacatagccccaaacttttttt
cggtggccagaaaatcccgtctaaggctgaacaaaaatagtccatctatctttgcatctccaactctatcggcatgtgtagcttcatagtttaa
agaacagagcaaaaaaaaaaaactaaagttttcgaaatttagcgatattccattgaaatatacatatgacattttgtattaacaattgcat
acatgcatttttgaaaaacatcttagaattaccgaaagatcgcaatgtcatcgattcaatattctttcaattcgttttaaagtttgccgactt
tatctcagtgctatttgtttagtgctgtcttgggaagtttgctgatatacatatgcatgtatgtatttttgtt**ggtcgccagcgctaaaggtca**
**tgctcttggcgatcagaaagagacggcagaacggcgaaagaaagagacgcccag**atgccaagtgacagacgtgagtgacagagagcaaacgaaa
aaaaatacagaaatgaaaagcgcaagaaaattctgttatccgaaaataaacaagaatattatctcggcacccagacgtggaaatattagttttg
gaaagccgaaga**GC**tg**GATTTT**cccgcttatc**GTG**c**GTGTGTTTAA**cg**GA**t**TTTCCC**cgac**GATAACCTCAAGTGC**a**ACATTGACGAGA**c**G**tcgg
gggtctgaagtggagaagtatctaagtatctaagtatcggtataggagtatgggagtagcaggtatagccagttcaatttccttcttccctggg
gtctttttttctattccattccaggggtatttcgaaaatatatacatacgtagttgggcttaggggtgaatatagttaaaaactcgagttttcaa
ttgaaattattaccaacttccaacagaactacaacaatttgcacgaagatcaggctgttttggtgttcttaattatttttttttctagattaatt
ctagctgaaatgattttaaaaggtatcccttagtcgaaccgaacgcatttggttcttagaaatatttttttcttttgttttttttggtgttgattg
ttttccccccgtggcttcaatttggaactcctcaagtttcccctgtcgagtgggtccttttttctctttatttttgtgccctgtgccacttgtgtgt
gcaccacttttcgggcacttttccccagagatcacaacaagtgcacgagtgtccgtccagggaatggggaaatttggtattttccctaacgaat
tccatccgtgaagcttcaaaattaatcagcgaaaaaaaataggatactccgtgaaag

**Figure 2.1.1.a. EvoDifference analysis of the distal 8-kb upper strand of the 40-kb *D. melanogaster dmyc* locus detects conserved sequence blocks.** The distal 8-kb fragment is named fragment D (see also Figure 2.2.a). Each block of uppercase bases represents a cluster of conserved sequences. Black capital letter bases are conserved in all species or in all but one species including *D. sechellia, D. yakuba, D. erecta, D. ananassae, D. pseudoobscura, D. willistoni or D. virilis*. There is one E-Box, conserved in all species except *D. virilis* (white text with black highlight). The neural network genetic algorithm PROMOTER 2.0 predicts three possible promoter regions using ccaat or bHLH recognition motifs (lower case black letters). The yellow highlighted sequence is predicted as a highly likely promoter region (dubbed as P0 putative). Sequence orientation is 5' → 3'.

**C**hrX:3,263,512-3,263,983

agtgtttcggcaacatttgagtttgcttggccacagtgtttctgcaacaaaaagaggggaaacttaaaaaaaactcactcatttcgttctgtga
tacactgtatatgaggcttttgcttttgcattgttttgatttttttcaatctagtgtgccgttataaaatcgccacgctgcacttttgttttgc
catttacgctcttggcgaatt**TGTTGTGTGA**gca**TGTGA**gtgtgtgtgtgt**GTGTGTGTGAG**cgcaactgtaccggcgtcttctttcttcgtcc
ttttttctttcgtgacgaagctttt**TTGTTG**t**TTTTG**c**AGCTG**c**C**g**GTTT**tcgccctttctattcaccttccaacatcttatgtctgtcttcacact
tacattctaggcgtatatttggaagtgcacatgcaaaacggcaaaaaaaatcgagagcgaagataaagacgaagaagacaacgaaataatttgt
aa

**C**hrX:3,264,456-3,264,691

atgacatgaatccatccatacaatcaaaatacctgtaaaactttccttattgcgtattataataaaagtacctacaattcgaaataacacttaa
agagagggtgtgatttttcgtcgccttagatttgaacgtatttttacgtttttgtagtaaaataccaatacatgggtaacgtgaaaaaaaat**AA**
**GAAG**tcga**AGATA**gcg**AC**a**T**tg**TTGCAA**aag**A**a**A**gactagtgcgcaat

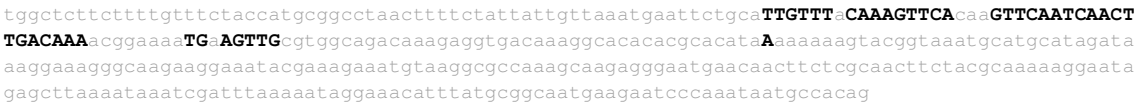
**C**hrX:3,265,046-3,265,281

tcccataactgccggtttattcgatttactgccaatgcattgcatttttacggccaatttttttgaaagcttgcaaattcttgacaaactagata
tggaaactcggcagccacttgctgccg**TTATTTACG**c**AGTTG**t**TTTT**atctcgccatgcacactttccatttacccgcgaaacaaaaaaaaaagtg
cagaaaatagttt**A**a**TA**ctatt**ACAA**a**AACA**agcagcacacgcaaaca

**C**hrX:3,265,636-3,265,753

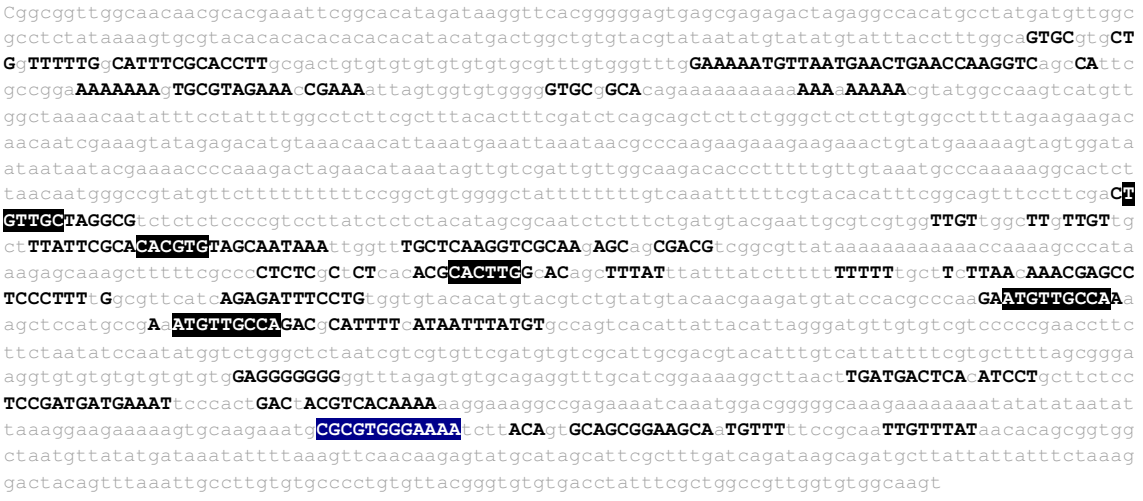
ctc**GTTGC**g**AAATCAAAACAAA**gcgtgtgaaagagagggcagatgcacgcgccaaaagcgctatgttgtgcactaaaagttgtggaaatgaaaa
tcaaagcgaaacgtaacaaaatcg

**C**hrX:3,266,226-3,266,579

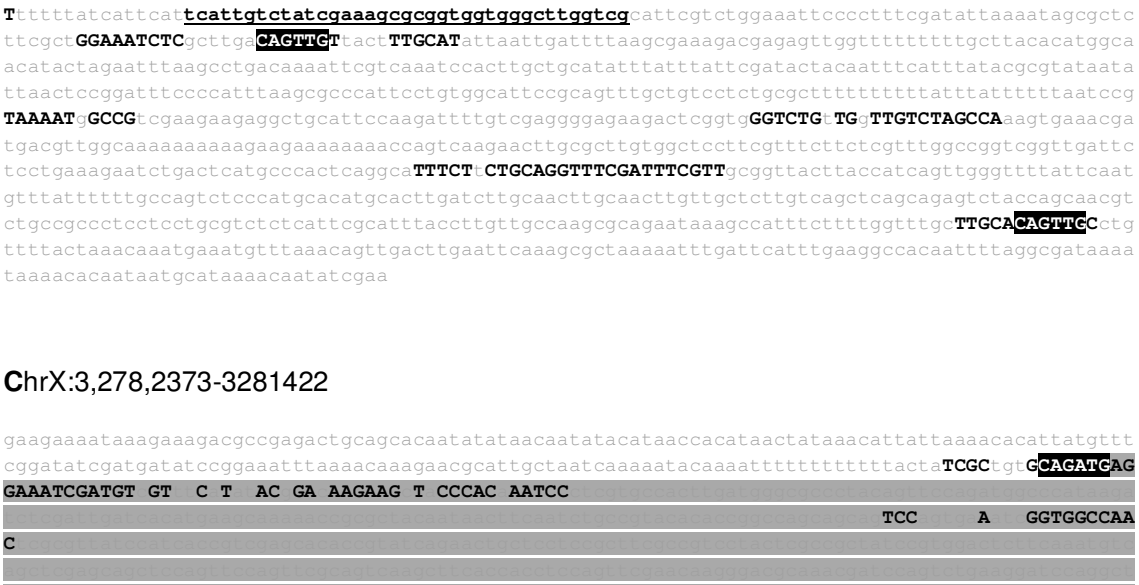


**C**hrX:3,269,878-3,270,827

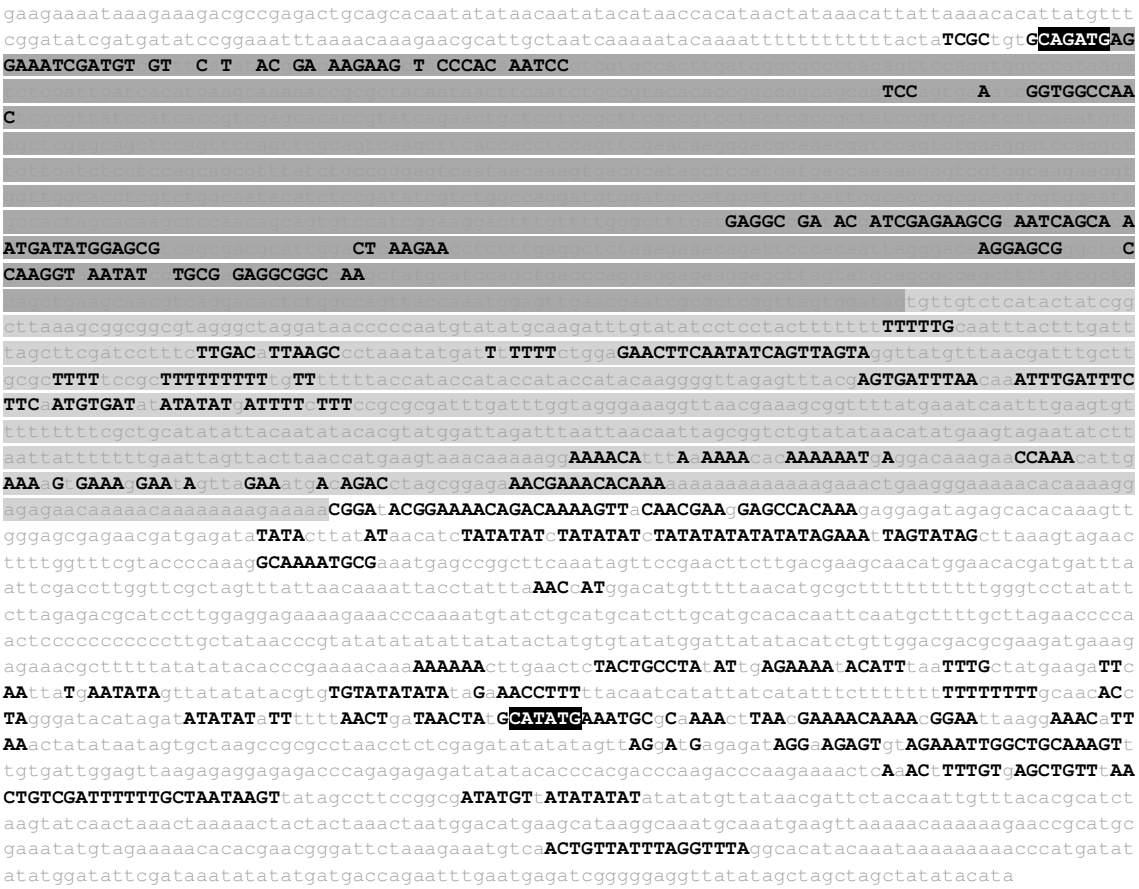
**C**hrX:3,272,691-3,274,460



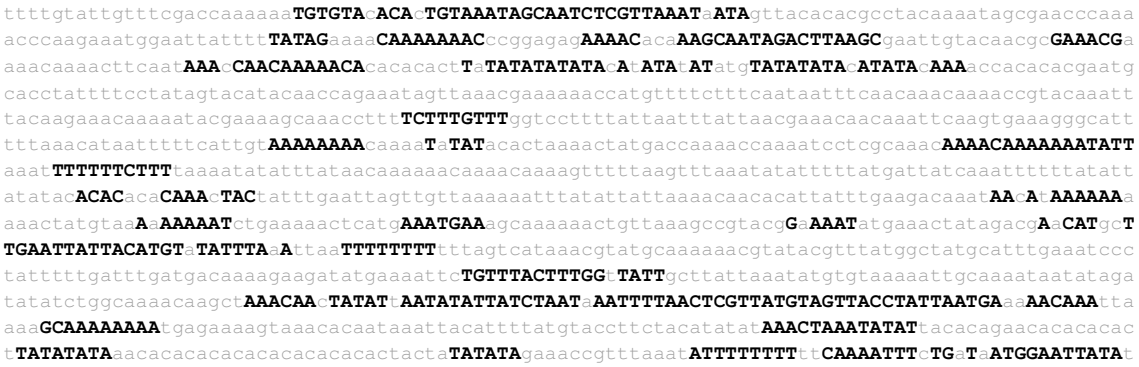
**C**hrX:3,275,523-3,277,528

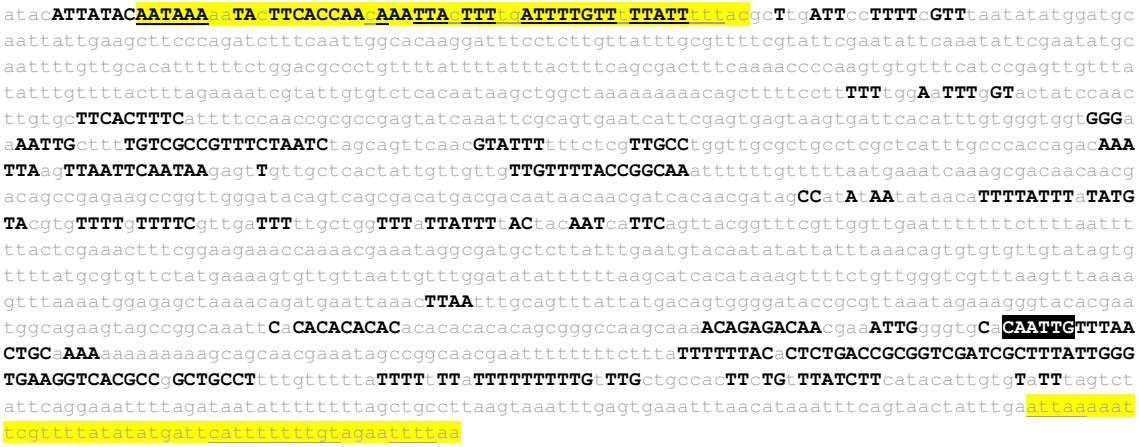


**C**hrX:3,278,2373-3281422

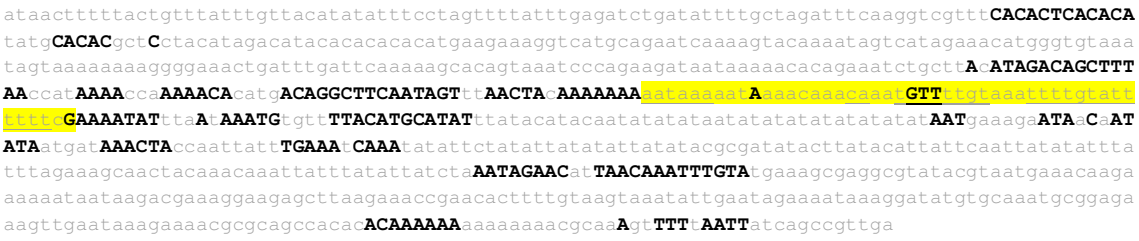


**C**hrX:3,281,777-3,284,608





**C**hrX:3,284,960-3,285,785



**C**hrX:3,286,494-3,286,847

**Figure 2.1.1.b. EvoDifference analysis of the 40-kb top strand *D. melanogaster dmyc* locus.** The 12.834-kb *dmyc* gene resides within the 40-kb fragment. Each block of uppercase bases represents a cluster of conserved sequences. Uppercase black nucleotides represent bases in the *D. melanogaster dm* EvoP 40-kb reference sequence that are conserved in all or all but one of the other 5 orthologous DNAs, *D. sechellia, D. yakuba, D. erecta, D. willistoni or D. virilis*. The *D. melanogaster dmyc* transcribed sequence is annotated according to FlyBase as follows: 5´-UTR (light grey), protein-coding sequence (dark grey) and 3'-UTR (light grey). The conserved bHLH binding sites and E-boxes are in white text with black highlight, as is the repeat motifs ATGTTGCCA and the nearby sub-repeat TGTTGC. The cluster CGCGTGGGAAAA, a dead-ringer for the HLHm-3-2 enhancer, is in white text with dark blue highlight. The *suppressor of Hairy wing su(H)* consensus binding site GTGGGAA is in the cluster. Poly (A1), (A2), (A3) in the 3'-UTR (yellow highlights), detected by Lasergene 9.1 GeneQuest module, is indicated. Regulatory elements within Poly (A1), (A2), and (A3) are underlined. The DPE in the intron 2 region is lower case bold and underlined. ORF orientation is 5' → 3'.

## 2.2 Source of the Genomic Sequences for the Creation of the *dmyc* Reporter Constructs

Two *Drosophila* bacterial artificial chromosomes (BAC) clones, Gene Bank accession numbers RP98-34B12 and RP98-2A13, contain genomic sequences of the *Drosophila* X-Chromosome. The 5'-end of the genomic fragment in the BAC clone RP98-34B12 contains the 8-kb far upstream region of the *dmyc* gene (Frag D in Figure 2.2.a). The BAC clone RP98-34B12 served as starting material to obtain the 8-kb fragment for the construction of JD transgene (See Materials and Methods, section 4.4.2) The BAC clone RP98-2A13 harbors 27-kb DNA sequences from the *dmyc* locus, including the gene itself (Figure 2.2.b). This BAC clone was used as substrate to clone the 27-kb *dmyc* locus (See Materials and Methods, section 4.3) (Figure 2.2.c).

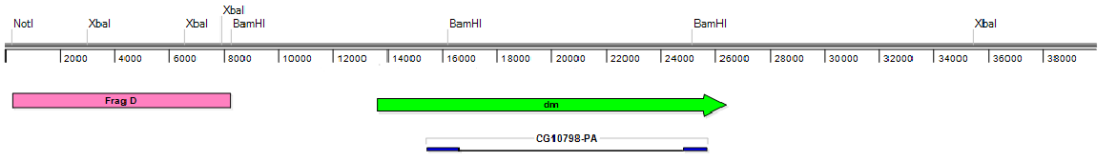



**Figure 2.2.a. Far upstream fragment (fragment D) and its relative distance to *dmyc* are shown.** Fragment D (8-kb in size) stems from the BAC clone RP98-34B12, and the 32-kb sequences adjacent to the fragment D is derived from the BAC clone RP98-2A13. The scale bar above the fragment D and *dmyc* gene represents the 40-kb genomic sequences from which all the deletion constructs were made. CG10798-PA = coding exons in *dmyc* protein. Scale bar is indicated in base pairs.

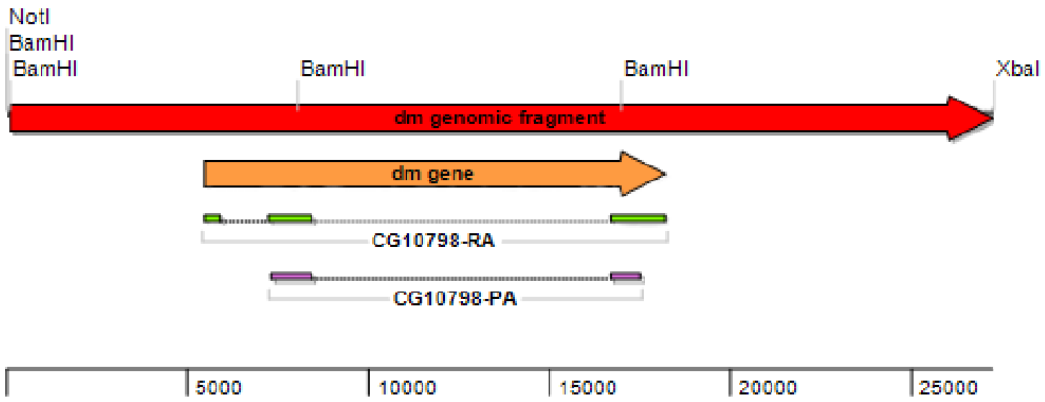



**Figure 2.2.b. 27-kb genomic sequences containing the *dmyc* gene.** The 5' NotI-XbaI 3' red arrow indicates the genomic sequences harboring the *dmyc* gene. The orange arrow represents the 12.8-kb *dmyc* gene. CG10798-RA = *dmyc* full-length RNA transcript; CG10798-PA = coding exons in *dmyc* protein. Scale bar is indicated in base pairs.

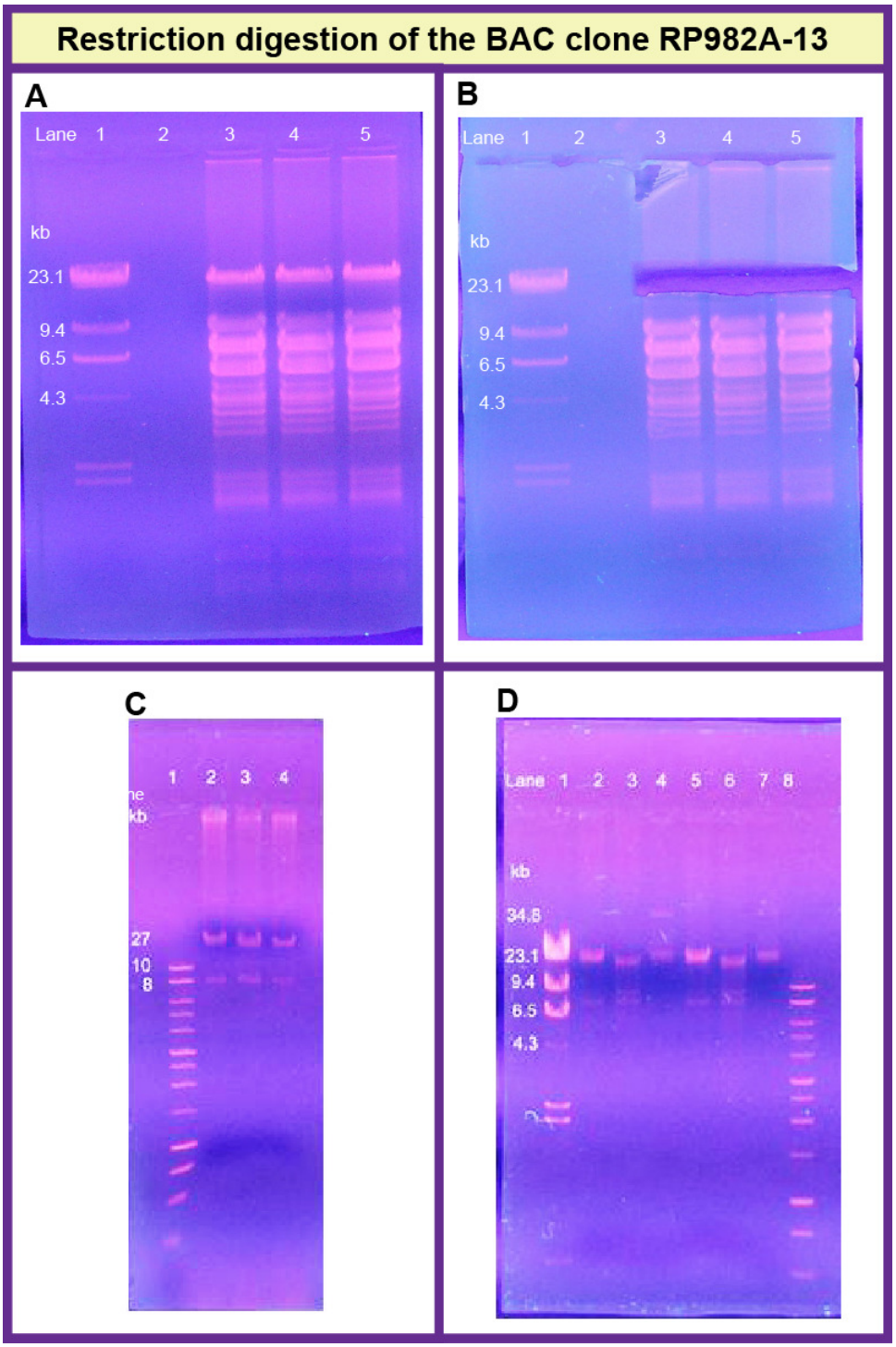



**Figure 2.2.c. Cloning of the *dmyc* gene from the BAC clone RP98-2A13.** The gene was cloned by triple digestion of *Drosophila* BAC clone RP98-2A13 with the enzymes NotI/DraIII/XbaI. (**A**) Lane 1, DNA marker HindIII; Lanes 3-5, ~27-kb NotI/XbaI fragment containing the *dmyc* gene. (**B**) The same as A indicating the gel cut of ~27-kb fragment. (**C** & **D**) Test digests of pC-RP27 with NotI/XbaI (C) or BamHI (D) resulting in RP27 genomic fragment of ~27-kb and vector 7.8-kb.

## 2.3 Experimental Identification of *cis*-Regulatory Regions within the 40-kb *dmyc* Locus

Based on computational analyses for the existence of conserved sequence blocks throughout the noncoding regions of the 40-kb *dmyc* locus, I synthesized a diverse array of reporter constructs from the *dmyc* promoter and other potential regulatory regions. In the reporter constructs the expression of a β-galactosidase cDNA is under control of different fragments of the noncoding regions. The sequence from the far upstream region contains a fragment of 8-kb in size (Frag D in Figure 2.2.a). The sequence from the proximal region of the *dmyc* gene includes approximately 5.2-kb upstream of exon1, exon1, intron 1, and the untranslated region of exon 2 (see section 2.4.2). The intragenic sequence of the *dmyc* gene is derived from the full-length large intron, from which different truncations have been produced (see section 2.4.3). To test the genomic fragments for *cis*-regulatory activity, these fragments were subcloned either into pCaSpeR4 vector for random P-element transformation or into phage ΦC31 pattP transforming vector for site specific mutagenesis. Both vectors contain a mini white gene that enables the identification of transgenic flies based on their eye color. Furthermore, the full-length *lacZ* cDNA (~ 3.1-kb) and SV40 polyadenylation signal (~ 0.7-kb) have been subcloned into the same vectors. Those constructs that are used to examine the *dmyc* 3'-UTR, contain either the largest fragment from the *dmyc* 3'-end (10.3-kb) or its deletions as transcription termination signal (see section 2.4.2 and Materials and Methods for subcloning details).

The resulting transgenes were injected either into $y^1$ $w^{1118}$ fly embryos or into embryos taken from different attB fly strains (Table 2.3) to establish independent transgenic fly strains. Expression of the *lacZ* reporter in larval tissues, embryos and ovaries under the control of *myc* genomic sequences was compared with the *dmyc* patterning in the same developmental tissues taken from the wild type *dmyc lacZ* enhancer trap line $w^{67c23}$ *P{lacW}*$dm^{G0354}$*/FM7c* (see Results, section 2.5.1). In all the experiments the fly stock "Blue balancer" and the enhancer trap line $y^1$ $w^{1118}$*; dpp-lacZ* were used as a source for positive control in larval brains and discs, the embryo and ovary staining. For the negative controls, dissected brains and imaginal tissues, embryos, and ovarian tissues from $y^1$ $w^{1118}$ and different attB fly lines were used (Table 2.3). The results of positive and negative staining reactions are given in the Appendix, section 5.2.

| NO | Bloomington Stock # | Name (Shorthand) | Genotype with FlyBase Links |
|---|---|---|---|
| 1 | 6598 | y w | *y*[1] *w*[1118] |
| 2 | 23648 | attp-86F | *P{hsp70-flp}1*, *y*[1] *w*[*]; *M{3xP3-RFP.attP}ZH-86Fb*; *M{vas-int.B}ZH-102D* |
| 3 | 24480 | attp-2A | *y*[1] *M{3xP3-RFP.attP}ZH-2A w*[*]; *M{vas-int.Dm}ZH-102D* |
| 4 | 24485 | attp-68E | *y*[1] *M{vas-int.Dm}ZH-2A w*[*]; *M{3xP3-RFP.attP'}ZH-68E* |
| 5 | 8622 | attp-8622 | *y*[1] *w*[67c23]; *P{CaryP}attP2* |
| 6 | 25709 | attp-25709 | *y*[1] *v*[1] *P{nos-phiC31\int.NLS}X*; *P{CaryP}attP40* |
| 7 | 25710 | attp-25710 | *y*[1] *sc*[1] *v*[1] *P{nos-phiC31\int.NLS}X*; *P{CaryP}attP2* |
| 8 | 24872 | attp-24872 | *y*[1] *M{vas-int.Dm}ZH-2A w*[*]; *PBac{y*[+]*-attP-3B}VK00037* |
| 9 | 11981 | dmyc-lacZ | *w*[67c23] *P{lacW}dm*[G0354]*/FM7c* |
| 10 | 2475 | double balancer | *w*[*]; *T(2;3)ap*[Xa], *ap*[Xa]*/CyO*; *TM3*, *Sb*[1] |
| 11 | 11108 | Blue Balancer | *Cyo, P{lArB}A66.2F2/b*[1] *Adh*[*] *cn*[*] *l(2)*[**]; *ry*[506] |
| 12 | 8412 | dpp-lacZ | *y*[1] *w*[1118]; *P{dpp-lacZ.Exel.2}3* |

**Table 2.3. List of the fly stocks used in the study.** For random P-element mediated transgene introduction embryos were taken from the *y[1] w[1118]* (Bloomington Stock number, 6598) flies, and for phi-C31 transgenesis embryos taken from different attp lines were used. The fly stock "Blue Balancer" that expresses *lacZ* in embryos and ovaries, was used as a source for positive control in the embryos and ovaries staining. The *dpp-lacZ* line was used as positive control for the imaginal discs staining.

## 2.4 Reporter Constructs Derived from the 40-kb *dmyc* Locus

In the following sections, I shall explain the designation of many different full-length or truncated sequences from the *dmyc* noncoding upstream, intragenic, and downstream regions that were used to study reporter activity under control of the *dmyc cis*-acting elements.

### 2.4.1 Reporter Construct JD

This construct contains the 8-kb genomic sequences from the *dmyc* far upstream region. To obtain the 8-kb genomic sequences for fragment D (see frag D, Figure 2.2.a), the BAC clone RP98-34B12 was amplified with the primer pairs PDF-1/PD-R1 and PDF-2/PD-R2 in two separate steps by polymerase chain reaction (see Appendix, Table 5.1). The amplified products D1 and D2 were combined by blunt ligation into the full-length fragment D (Figure 2.4.1). The 5' NotI-Acc65I 3' fragment D is fused to the *lacZ* reporter gene in the fly transforming vectors pCaSpeR4-NLSlacZ and pattB-temp del LoxP (Materials and Methods, section 4.4.2). The SV40 polyadenylation site has been used to terminate transcription.

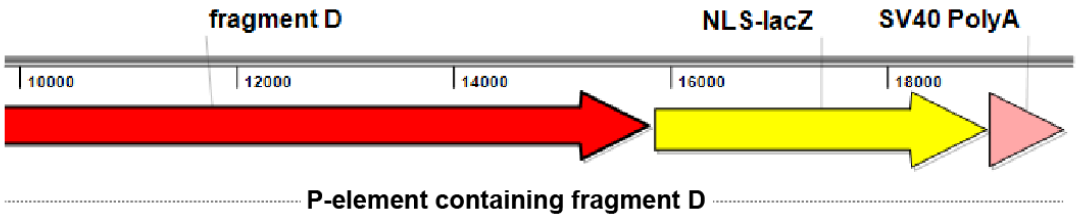



**Figure 2.4.1. P-element containing the far upstream fragment in the JD transgene.** The 8-kb far upstream fragment D has been subcloned in both fly transforming vectors, pCaSpeR4-NLSlacZ and pattB-NLSlacZ (see Materials and Methods, section 4.4.2) Scale bar is indicated in base pairs.

### 2.4.2 Reporter Constructs Derived from the *dmyc* Proximal Region

The largest fragment from the *dmyc* upstream noncoding proximal region is ~7.7-kb in size. Initially, 3'-end of a ~7.9-kb *dmyc* genomic fragment has been truncated to remove the ORF sequence of the second exon at the 3'-end of the sequence, but retain all the noncoding nucleotides (see Materials and Methods; section 4.4.1). The resulted ~7.7-kb 5' NotI-Asp718 3' fragment from the *dmyc* 5`end, in the constructs J2.1 and the J2.1 truncations, is fused to the *lacZ* reporter in which transcription terminates by the SV40 polyadenylation signal (Figure 2.4.2.a).

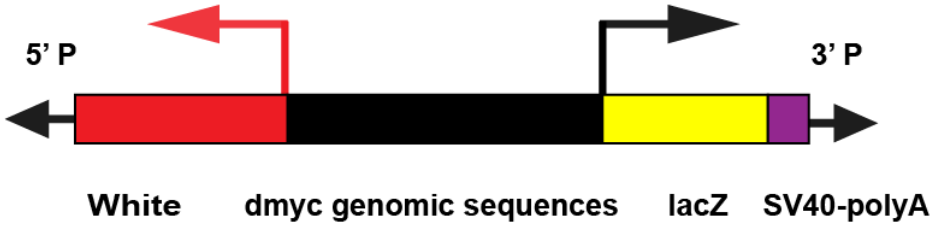



**Figure 2.4.2.a. Basic structure of the *dmyc-lacZ* transgenes derived from the *dmyc* far upstream, proximal, and intronic regions.** *dmyc* genomic sequences driving *lacZ* reporter transcription are the noncoding regions originating from far upstream fragment D, 5' sequences or the intron 2 region. Transcription termination occurred through the SV40 poly (A) signal.

The deletion constructs (J3-J7), derived from J2.1, have been made extending from minus 5.2-kb upstream of transcription start to minus 0.1-kb upstream of transcription start. All the constructs retain the transcription start, exon 1, intron 1, and the exact beginning, non-coding region of the exon 2 (Figure 2.4.2.b). The deletion constructs were made by considering the identified conserved sequence blocks and based on the availability of convenient restriction sites in these regions.

Like J2.1, the largest construct J1 contains the ~7.7-kb fragment from the *dmyc* 5' end for its promoter, but a 10.3-kb BamHI-XbaI (where the BamHI site is located 554 bp upstream of the *myc* translation stop and the XbaI site is 9763 bp downstream of the translation stop) fragment of the *dmyc* 3' region as polyadenylation signal (Figure 2.4.2.c). The reason for including this fragment was to determine whether any of the *cis*-elements are located 3' of the *dmyc* ORF. Furthermore, the aim was to identify potential poly-adenylation signals.

Derived from J1, a series of deletion constructs (J1.1, J1.3-J1.5) have been generated extending from nucleotide minus 554 upstream of the translation stop to nucleotide plus 9763 downstream of the translations stop (Figure 2.4.2.d). The deletion constructs retain poly (A1), poly (A2), poly (A3) or none of the predicted poly (A) signals, namely in the deletion J1.3 (see Results, section 2.5.6 and Discussion, section 3.1.5).

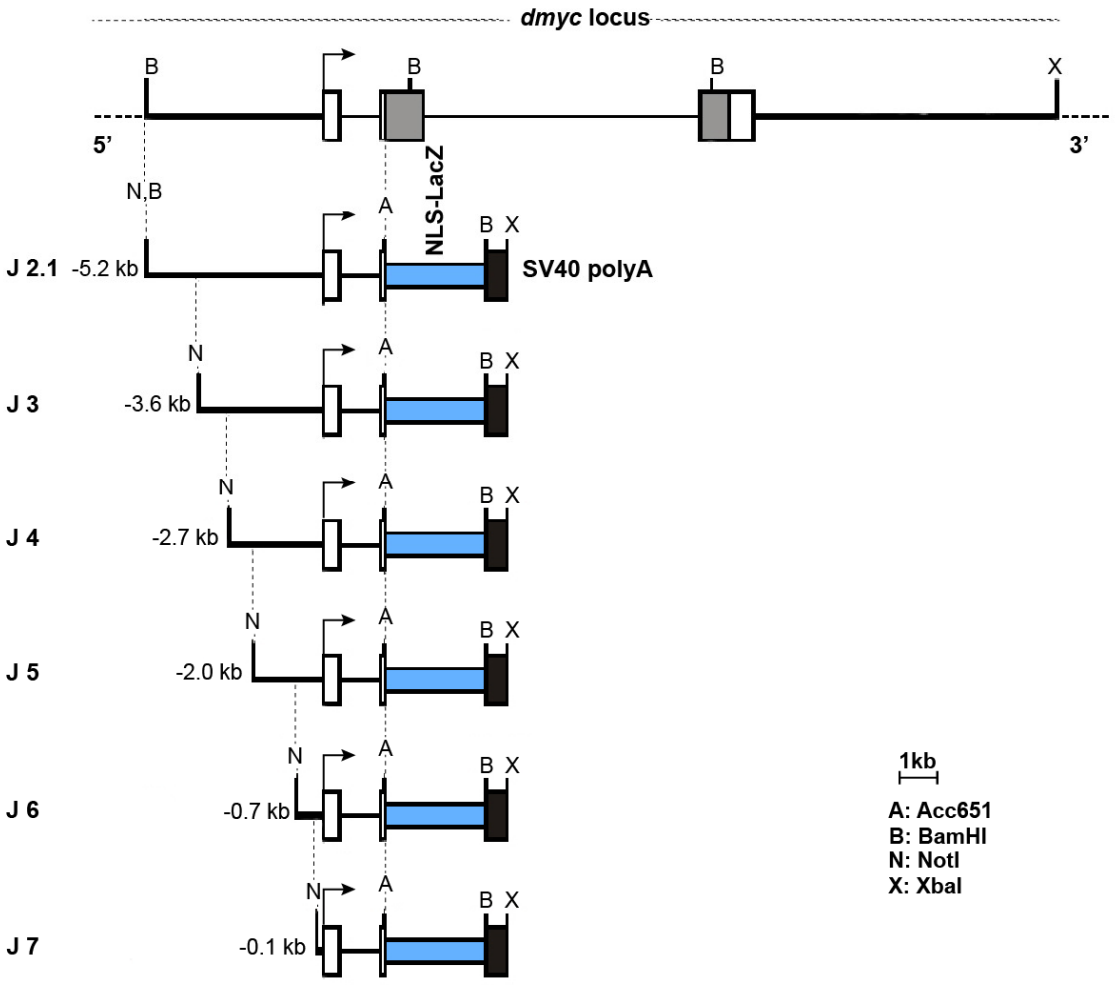



**Figure 2.4.2.b. *dmyc-lacZ* reporter constructs J2.1-J7.** In these constructs transcription of the reporter gene is under control of different sequences from the 5'-end of the *dmyc* gene. SV40-poly (A) terminates transcription of *lacZ*.

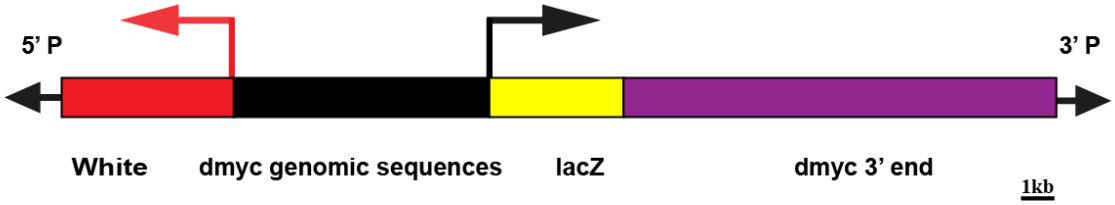



**Figure 2.4.2.c. Basic structure of the *dmyc-lacZ* transgenes containing the *dmyc* 3'-UTR sequences.** *dmyc* genomic sequences driving *lacZ* reporter transcription are the noncoding regions originating from the 5'-end sequences. *dmyc* full-length poly (A) or its truncations terminate transcription of *lacZ*.

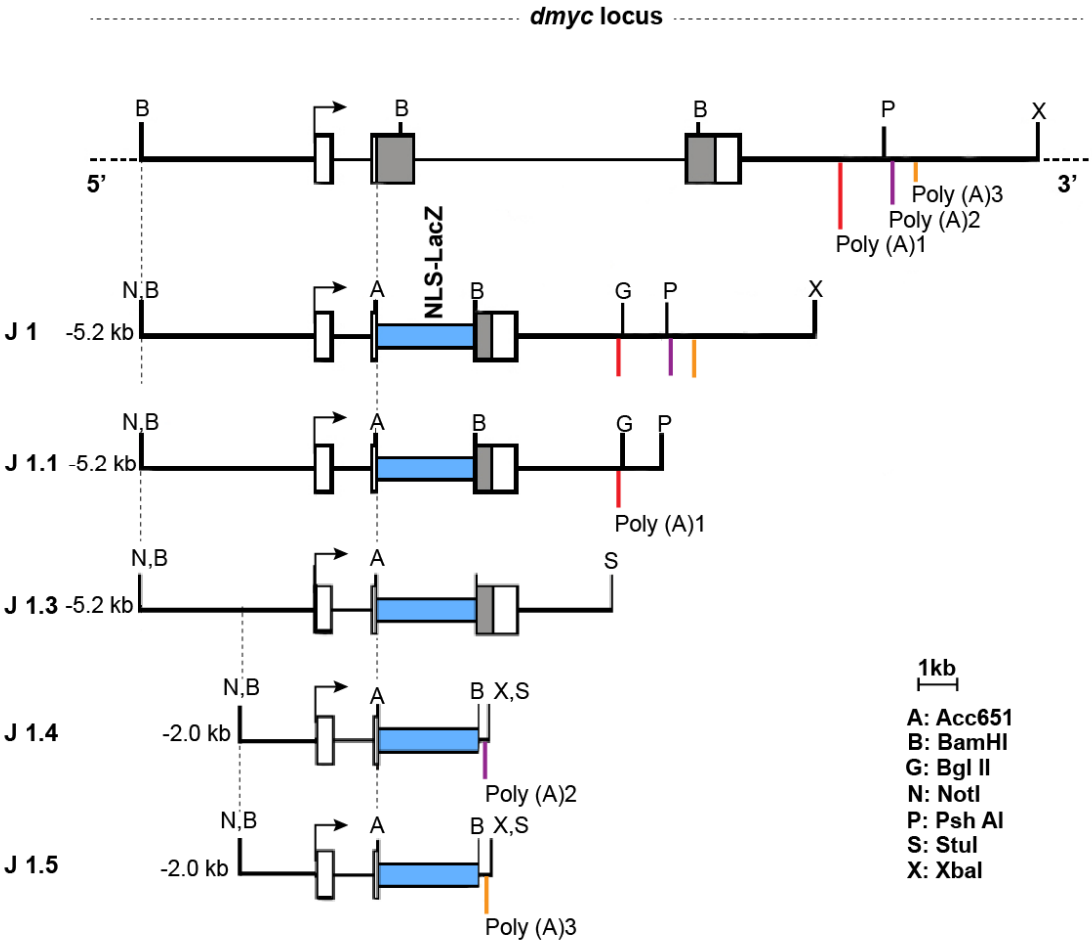


**Figure 2.4.2.d. *dmyc-lacZ* reporter constructs J1, J1.1, and J1.3-J1.5.** Like J2.1-J7 (Figure 2.4.2.b), transcription of the reporter gene is under control of sequences from the 5'-end of the *dmyc* gene. In contrast to J2.1-J7 constructs, *dmyc*-poly (A) terminates transcription of *lacZ*. Truncations of J1 at its 3' end include J1.1, J1.3, J1.4, and J1.5. In the J1 fragment, three potential poly (A) signals with a high degree of homology to the consensus polyadenylation signal were detected computationally. The transgenes J1, J1.1, and J1.3 are under control of the full-length 5'-promoter (see also J2.1 in Figure 2.4.2.b). The constructs J1.4 and J1.5 contain the promoter of J5 in Figure 2.4.b. J1.1 only contains poly (A)1, J1.3 does not include the predicted poly (A) sites, J1.4 contains poly (A)2, and J1.5 contains poly (A)3.

### 2.4.3 Reporter Constructs Including the *dmyc* Intron 2 Region

To analyze the intragenic region of the *dmyc*, the constructs J8 and its deletions J8.2, J8.4, and J8.5 (Figure 2.4.3.a) has been generated from this region. A BamHI-BamHI fragment of ~8.2-kb in size from the *dmyc* intron 2 region has been truncated by removing the ORF sequences at the 5' and the 3'-end of the sequence to generate intron 2 minus ORFs (intron 2 fragment free from coding sequences (see Materials and Methods, section 4.4.3). The constructs J8 and J8.1 both contain the ~8.1-kb 5' NotI-Acc65I 3' fragment from the *dmyc* intron 2 region minus ORFs as promoter, which is fused to the *lacZ* reporter (Figure 2.4.3.a). In order to obtain the intron 2 sequence for the creation of J8 and J8.1, the plasmid pC-RP27 (originated from the BAC clone RP98-2A13, see Materials and Methods, section 4.3) was used as substrate. The sequences in the plasmids J8.2, J8.4, and J8.5 are derived from the J8 construct. J8.1 is very similar to J8 except the orientation of intron 2 fragment is 3' $\rightarrow$ to 5' (Figure 2.4.3.a). In all constructs SV40 poly (A) signal terminates the transcription of *lacZ*.

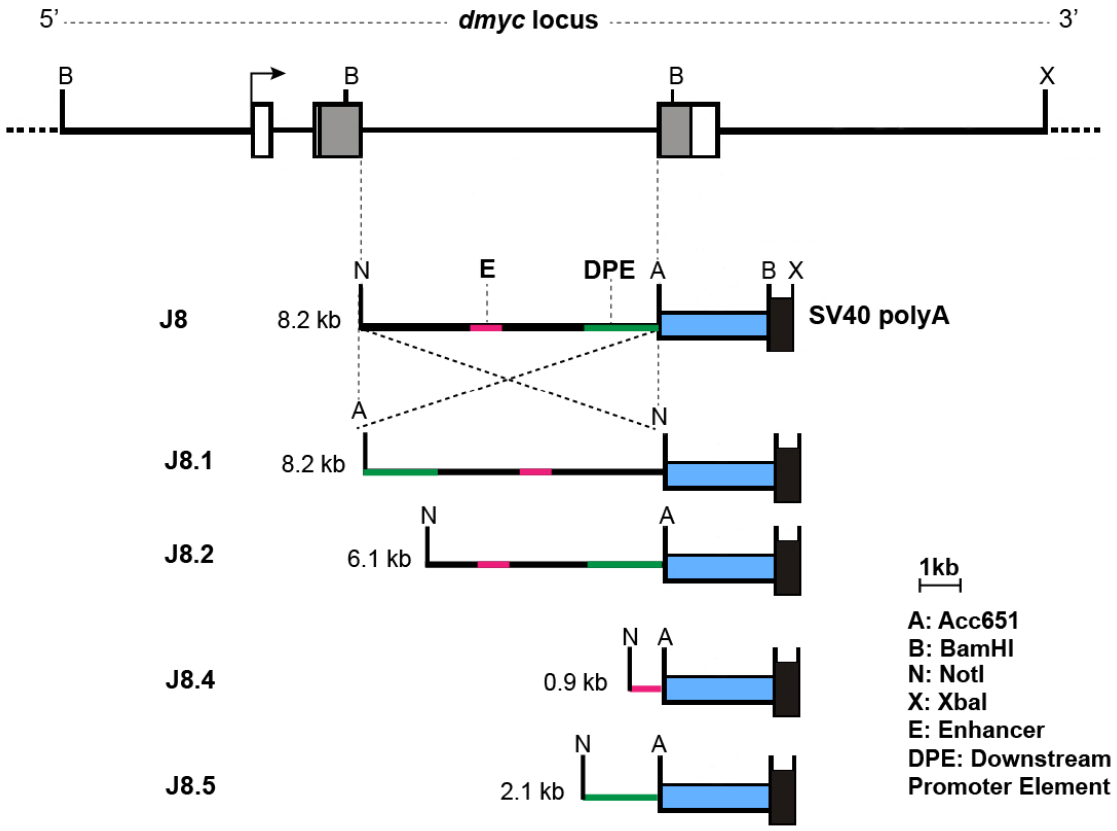



**Figure 2.4.3.a. *dmyc-lacZ* reporter construts J8, J8.1, J8.2, J8.4 and J8.5.** In these constructs transcription of the reporter gene is under control of different sequences from the *dmyc* intron 2 region. SV40-poly (A) terminates transcription of *lacZ*.

The logic of making these constructs was threefold: (i) to see if there are any downstream promoter elements (DPE) in the intron 2 sequence, as is the case in many developmentally switched-on/off genes in *Drosophila* (Burke et al., 1998); (ii) to see if the 2.5-kb sequences in J8.5 from the 3'-end of the intron that include the putative enhancer region and the sequences downstream of the enhancer region, would be sufficient for the reporter expression; (iii) the reverse orientation of the intron sequence in J8.1 may give some information about the directionality of regulation from this region.

The construct J9 contains a 5' NotI-Asp718 3' fragment including the *dmyc* intron 2 region minus ORFs, which is fused to the promoter of deletion construct J5 in front of

the *lacZ* reporter (Figure 2.4.3.b). The purpose of making this construct was due to the fact that the deletion J5 showed the same activity for *lacZ* as did the main construct J2.1. By fusing J5 promoter to intron 2 sequences, the aim was to obtain the most possible expression of the reporter gene mimicking the endogenous *dmyc* activity.

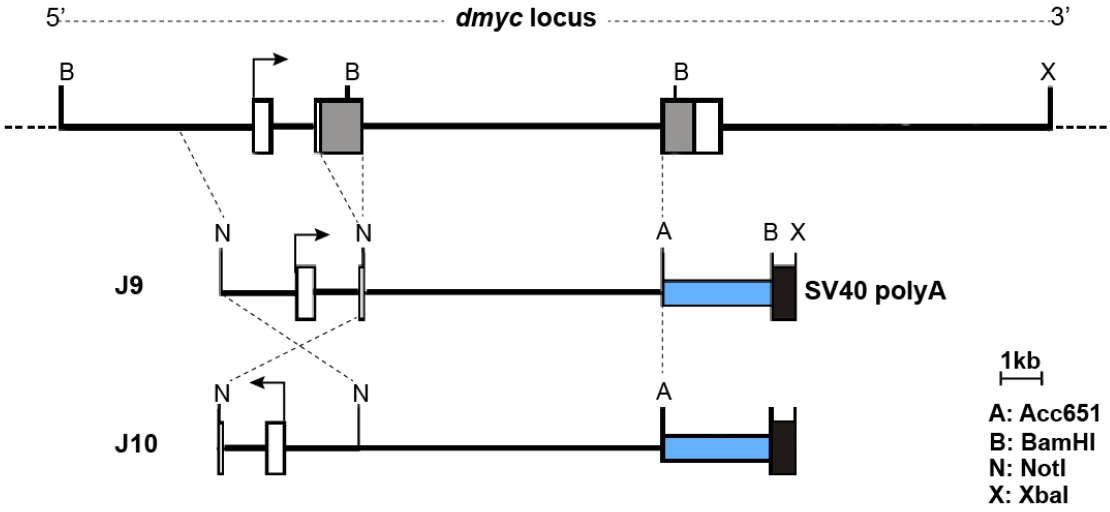



**Figure 2.4.3.b. *dmyc-lacZ* reporter constructs J9 and J10.** In these constructs transcription of the reporter gene is under control of sequences from the 5'-end of the *dmyc* gene fused inframe to intron 2 sequence (J9) or fused in reverse orientation to intron 2 sequence (J10). SV40-poly (A) terminates transcription of *lacZ*.

The construct J10 is very similar to J9 except the orientation of the J5 promoter is 3' → to 5' (Figure 2.4.3.b). The reason for creating this construct was to test whether parameters such as stoichiometry, affinity, spacing and arrangement of binding sites within the *cis*-regulatory regions influence the output of the *dmyc* transcriptional regulatory sequences. For example, depending on the total number of binding sites for a certain regulator in the same promoter region, stage-specific and tissue-specific expression patterns can be achieved for the same gene during development (Kulkarni and Arnosti, 2005).

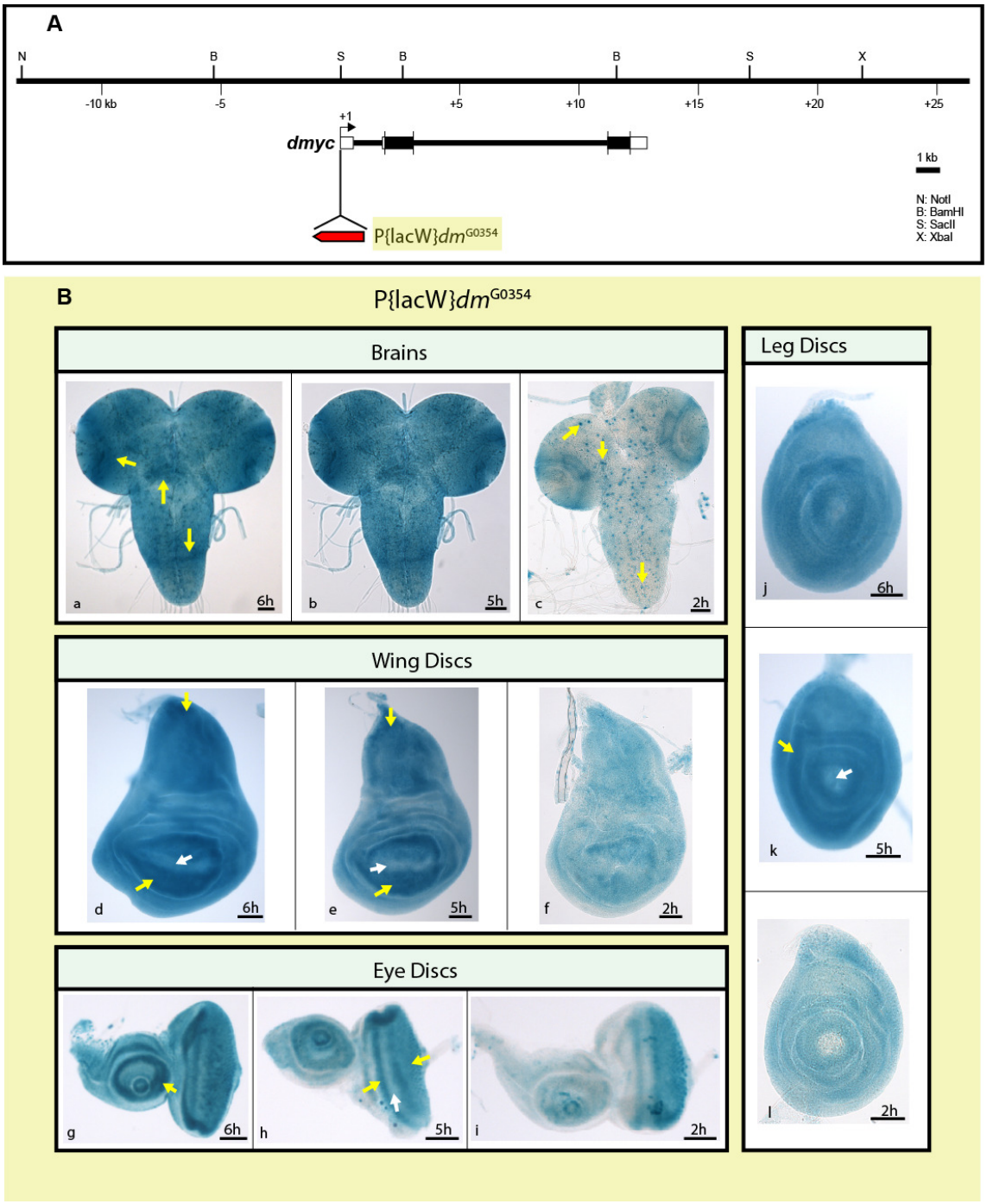
## 2.5 *dmyc-lacZ* Activity in the Larval Imaginal Tissues, Embryos, and Adult Female Tissues

In the following sections, the expression patterns of *lacZ* reporters driven by *dmyc* genomic sequences in the brain, and the wing, leg, eye-antennal imaginal tissues, embryos, and ovary will be described. Two principles of the analysis of the *lacZ* expression in each tissue are possible, "vertically", looking at each construct separately or "horizontally", comparing the expression of *lacZ* of all constructs in one tissue, for example only the leg discs.

For simplicity in comparing each fragment with its deletions from a specific region and their ability in controlling the reporter expression, I have presented the data in this section "horizontally" by showing one representative line for the activity of each construct in that region in all the tested tissues. In Appendix (section 5.4), I have extended the horizontal presentation by including more than one line per construct.

### 2.5.1 *lacZ* Expression in the *dmyc-lacZ* Enhancer Trap Line Similar to the Endogenous *dmyc* Patterning

I was interested in comparing the pattern of *dmyc* promoter activity from the different *dmyc* deletion constructs generated from upstream, intronic, and downstream regions, with both the endogenous patterning of *dmyc*, as previously implicated, and with the pattern resulting from the *dmyc-lacZ* enhancer trap line $w^{67c23}$ *P{lacW}dm*$^{G0354}$*/FM7c* (Figure 2.5.1.a/A). The enhancer trap line $w^{67c23}$ *P{lacW}dm*$^{G0354}$*/FM7c* has been established to give patterns of endogenous *dmyc* expression (Figure 2.5.1.a/B and Figure 2.5.1.b), and is known to be responsive to *dmyc* regulators (Cranna and Quinn, 2009; Cranna et al., 2011; Mitchell et al., 2010; Siddall et al., 2009).


A
N
B
S
B
B
S
X
-10 kb
-5
+5
+10
+15
+20
+25
+1
dmyc
1 kb
N: NotI
B: BamHI
S: SacII
X: XbaI
P{lacW}dm^G0354
B
P{lacW}dm^G0354
Brains
Leg Discs
Wing Discs
Eye Discs
a 6h
b 5h
c 2h
d 6h
e 5h
f 2h
g 6h
h 5h
i 2h
j 6h
k 5h
l 2h

Results

**Figure 2.5.1.a. The *dmyc-lacZ* enhancer trap line *w*$^{67c23}$ *P{lacW}dm*$^{G0354}$*/FM7c* reflects the endogenous *dmyc* patterning in the larval brain and imaginal tissues.** (**A**) Insertion site of the P-element containing reporter *lacZ* at the *dmyc* locus is shown. Breakpoints of the insertion are as follows: 3D2, X:3267141..3267197, which maps to the region 213 nucleotides upstream of *dmyc* exon 1 start site (Bloomington stock number 11981, donors, Ulrich Schaefer and Herbert Jackle). (**B**) Third instar larval brain and discs were assayed with β-gal reaction for *lacZ* expression. In all the tested imaginal discs and larval brain, *lacZ* patterning reflects the pattern reported for *dmyc* endogenous mRNA distribution (**a**–**c**, brain; **d**–**f**, wing discs; **g**–**i**, eye discs; **j**–**l**, leg discs). Yellow arrow indicates *lacZ* expression and white arrow indicates lack of *lacZ* activity. Details on *dmyc* patterning are explained in the Discussion section. Staining time is indicated above the scale bar. Scale bar in (**a**–**l**) indicates 50 µm.

β-gal assays for this *dmyc-lacZ* enhancer trap line revealed ubiquitous *dmyc* promoter activity in the larval brain, with an increased level of expression in the distal and middle parts of the lobes and within dividing neuroblasts in the middle parts of the ventral ganglion (Figure 2.5.1.a/B, a, yellow arrows). Additionally, *lacZ* activity is restricted to a limited number of cells distributed in the two proximal halves of the hemispheres and along the ventral ganglion (Figure 2.5.1.a/B, c, yellow arrows). The enhancer trap line detected *lacZ* activity around the wing pouch and in the notum region (Figure 2.5.1.a/B, d–f), anterior and posterior to the morphogenetic furrow in the eye disc and around the center of the antennal disc (Figure 2.5.1.a/B, g–i), and in concentric rings around the center of the leg disc (Figure 2.5.1.a/B, j–l). dMyc antibody staining of the tissues taken from the above enhancer trap line results in the same pattern as observed for the endogenous reporter (L. Quinn, personal communication).

During early stages of embryogenesis ubiquitous expression of *lacZ* was detectable (Figure 2.5.1.b/a). In later stages, β-gal assays for the *dmyc-lacZ* enhancer trap line revealed *dmyc* promoter activity mainly in presumptive mesoderm along the ventral midline. During germ band extension, *lacZ* expression is intensified in the mesoderm, midgut, pharynx, and anal pad (Figure 2.5.1.b/d-l). β-gal assays in the ovaries showed reporter activity mainly in the nurse cells and oocyte, but weakly expressed at the tip of germarium (Figure 2.5.1.b/m, n).

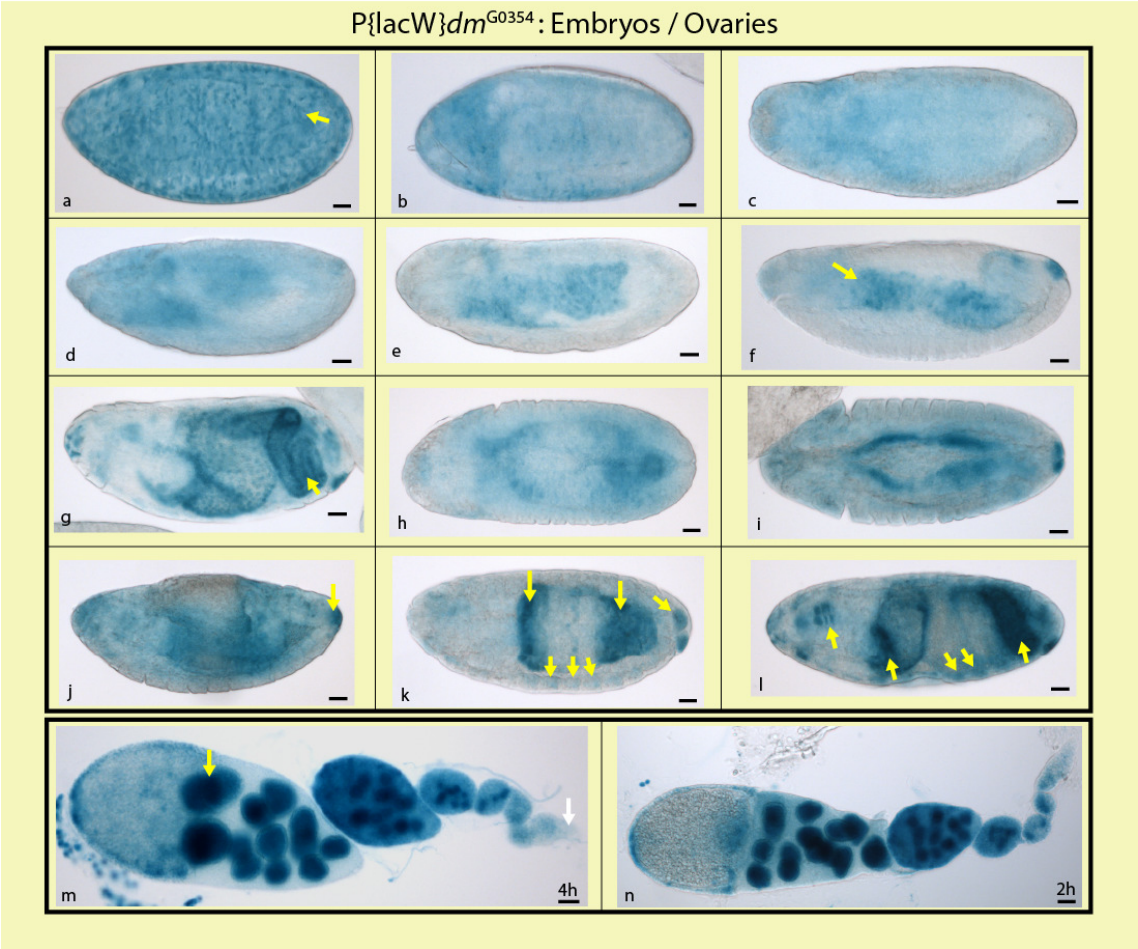



**Figure 2.5.1.b. The *dmyc-lacZ* enhancer trap line *w*$^{67c23}$ *P{lacW}dm*$^{G0354}$*/FM7c* produces the endogenous *dmyc* mRNA pattern during embryogenesis and in adult ovaries.** In developing embryonic tissues, the *dmyc-lacZ* enhancer trap expresses *lacZ* with high similarities to *dmyc* mRNA distribution, predominantly in the midgut, hindgut, pharynx, anal pad, and partly in the mesodermal tissues. **A**–**l** embryos (embryo stages are as follows: **a**, **b**, stage 1–4; **c**, **d**, stage 6–8; **e**, **f**, stage 9–12; **g**–**l**, stage 13–16). The staining in the ovaries (**m**, **n**) reflects *dmyc* mRNA localization, which has been reported in nurse cells, but is weakly expressed at the tip of the germarium. Yellow arrow indicates *lacZ* expression and white arrow indicates lack of *lacZ* activity. Staining time for ovaries is indicated above the scale bar, embryo staining took place over-night. Scale bar in (**a**–**n**) indicates 50 µm.

### 2.5.2 Reporter Activity under Control of *dmyc* far Upstream Fragment

Less than 5% of human *c-myc* transcripts initiate from the far upstream TATA less P0 promoter (Bentley and Groudine, 1986a, b). Initially, we had identified multiple conserved sequence blocks, including a conserved E-box, in the *dmyc* far upstream region (Figure 2.1.1.a). I decided to first test the 8-kb far upstream fragment for its ability to self-initiate transcription (Figure 2.5.2.a).

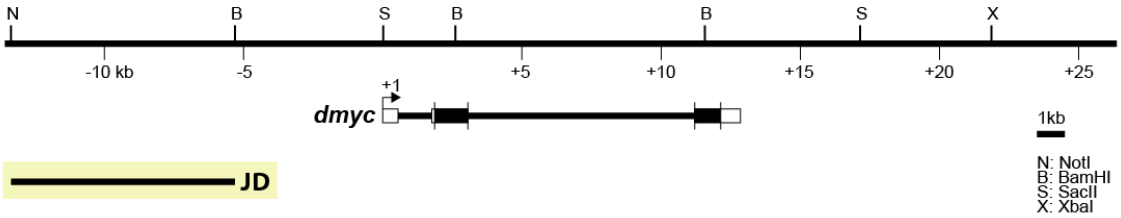



**Figure 2.5.2.a. Distal upstream regulatory region fragment D.** Locations of the *dmyc* and the transgene JD (8-kb in size) relative to the 40-kb genomic locus are indicated.

**JD**: for this transgene eight independent transgenic lines were analyzed, namely, JD-13.1, JD-42.2, JD-47.2, JD-51.1, JD-58.2, JD-61.4, JD-68E-21.1, and JD- 68E-26.

In all eight lines no activity was detected in the tissues known to normally express *dmyc*, including larval brain and imaginal tissues (Figure 2.5.2.b/A, a–e), the earlier embryonic stages (Figure 2.5.2.b/B, f, g), and adult ovaries (Figure 2.5.2.b/C, k, l). To confirm the absence of *dmyc* activity in the examined tissues, I performed reverse transcription polymerase chain reaction on the RNA extracts taken from the discs shown in Figure 2.5.2.b (c-e), and did not detect any *lacZ* transcripts beyond background level in these tissues that were negative for lacZ staining (Figure 2.5.2.b/D lanes –RT to e). Activity of the *JD-lacZ* reporter was confined to presumptive mesodermal tissues in body segments and the head regions of embryos during late development (Figure 2.5.2.b/B, h–j, arrow heads). Previous studies have reported *dmyc* activity in putative neuromuscular tissues by *in situ* hybridization experiments on *dmyc* endogenous mRNA. However, the enhancer trap line used as the control in this study shows only partial expression in these tissues (Figure 2.5.1.b/k, l). This result suggests that the regulatory sequences in this far upstream region are only sufficient for activation of *dmyc* expression during late embryogenesis. Due to its far upstream position, we have dubbed the remote *cis*-regulatory elements in this region “P0 (putative)”, analogous to the human *c-myc* P0 promoter.

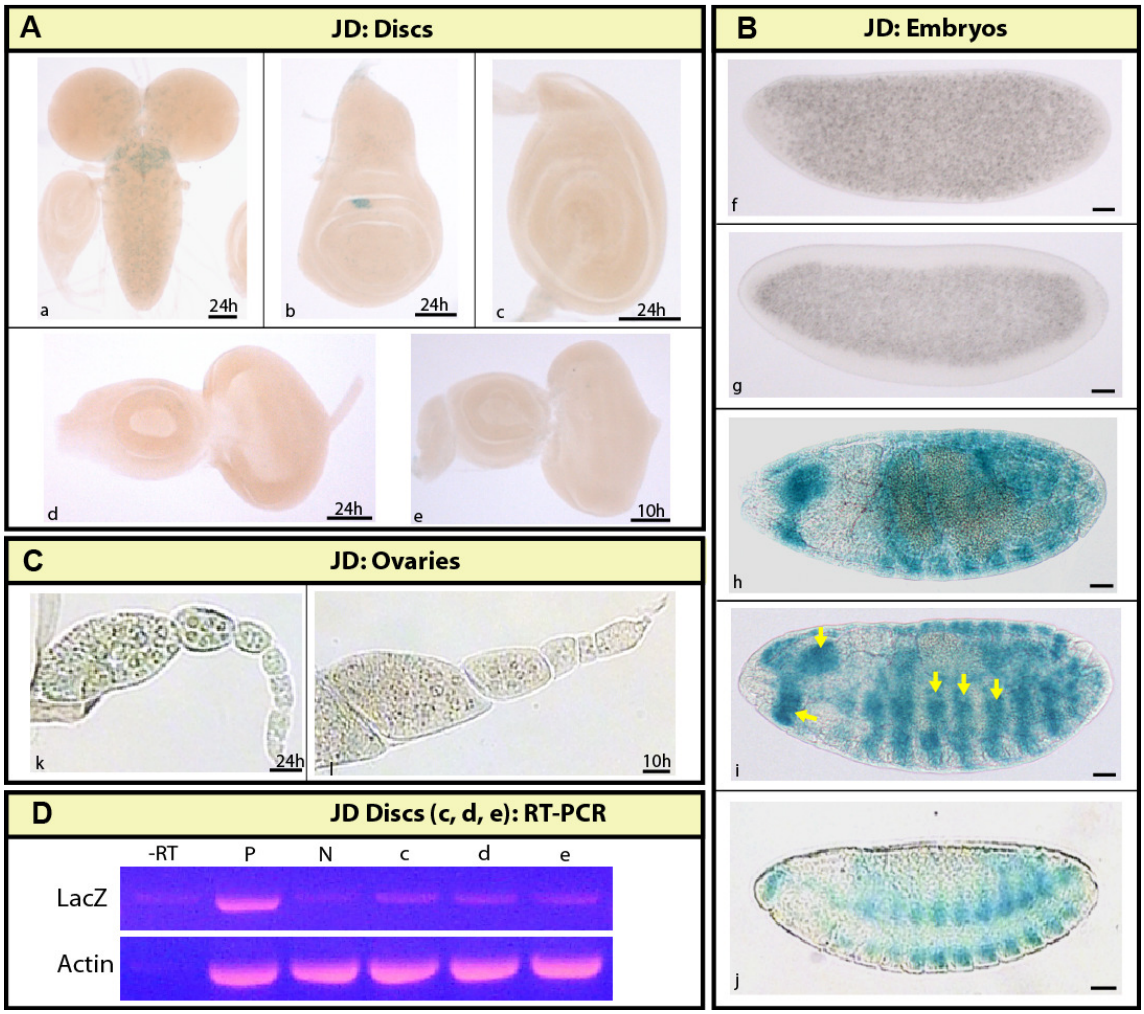



**Figure 2.5.2.b. Distal upstream regulatory region of *dmyc* is active in late embryogenesis.** (**A**) Activity of the JD transgene in the brain and imaginal tissues. The fragment was neither active in the brain (**a**) nor in imaginal discs (**b**, wing; **c**, leg; **d**, **e**, eye discs). (**B**) JD fragment was not active in early embryos (**f**, **g**), however, the fragment recapitulated expression of *dmyc* in late embryogenesis (**h**–**j**, stage 13–16). Yellow arrow indicates *lacZ* expression. (**C**) The staining in ovaries (**k**, **l**) showed no activity for the JD fragment. (**D**) Reverse transcriptase polymerase chain reaction on JD discs (**c**, **d**, **e** shown in **A**), detects no *lacZ* transcripts beyond background level. Controls used are as follows: -RT, negative control with no transcriptase (brains from **2.5.3.c**/**A, b**); P, **2.5.3.c**/**A, b**, positive control; N, *y[1] w[1118]* leg discs, negative control; actin, *Drosophila* actin as internal control. The tissues represented were taken from the transgenic line JD-13.1, except the eye disc in (**e**) is dissected from the stock JD-61.4. For more staining in JD transgenic lines see also Figure 5.4.a (Appendix, section 5.4). Staining time for discs and ovaries were either 10 or 24 hours, as indicated (**e**, shown in **A**, one representative for 10-hour discs). The embryos were stained over-night. Scale bar in (**a**–**j**) indicates 50 µm.

### 2.5.3 Activity of Proximal Upstream Region in Larval and Adult Female Tissues

The two main promoters required for activation of mammalian *c-myc* transcription, P1 and P2, are located in the 5'-UTR, where the majority of transcripts are initiated (Bareket-Samish et al., 2000; Stewart et al., 1984; Yean and Gralla, 1997). Thus, my first efforts were directed toward identifying regulatory elements in the *dmyc* 5'-UTR, capable of activating endogenous patterns of *dmyc* gene expression during development (Figure 2.4.2.b, Figure 2.5.3.a).

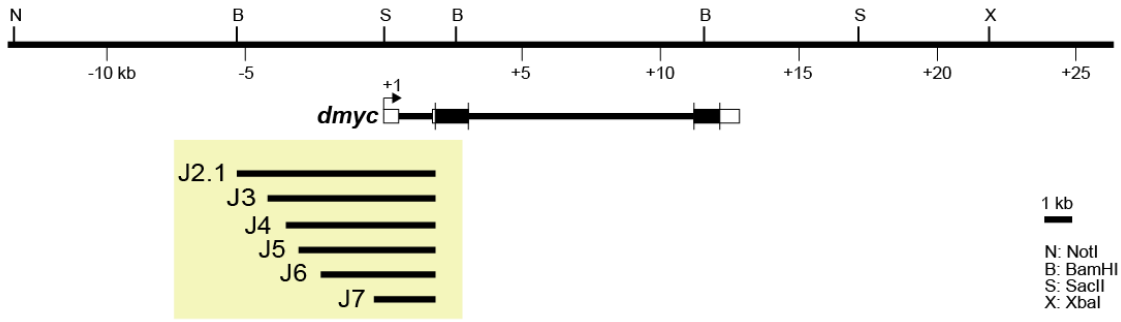



**Figure 2.5.3.a. Deletion constructs generated from the *dmyc* 5' regulatory region are shown.** The 40-kb genomic locus and the location of the *dmyc* is shown at the top. A fragment of 7.159-kb in size (J2.1, lacking open reading frame sequences) from the *dmyc* 5' region and its deletion restriction fragments (J3-J7) are shown below. See also Figure 2.4.2.b.

For each of the generated constructs from this region established independent transgenic lines were analyzed.

**J2.1**: for this construct the five independent transformants, J2.1-12.1, J2.1-30.1, J2.1-86F-1, J2.1-86F-7, and J2.1-86F-9 were looked at. There is a faint ubiquitous activity in the larval brain of two analyzed lines, J2.1-12.1 (Appendix, section 5.4), J2.1-30 (Figure 2.5.3.b/a), with an increased level of expression in the middle of the front part of both hemispheres. Also the middle part of the both sides of the ventral ganglion shows strong *LacZ* activity. The increased activity in the middle part of proximal and in the middle of distal parts of the hemispheres is much more intensified in the line J2.1-12.1. In addition, the staining in the brain of third instar larvae is restricted to a limited number of cells distributed in the two proximal halves of the hemispheres and along the ventral ganglion. The expression pattern observed in the brain shows high

consistency in terms of the number of cells staining for lacZ in all of the analyzed transgenic lines (Figure 2.5.3.b/a, Appendix Figure 5.4.b/a, b).

All five lines show an intermediate staining around the pouch in the wing discs, however, in the notum region the staining observed for the lines J2.1-30.1 and J2.1-86F-1 is weak (Figure 2.5.3.b/d, Figure 5.4.b/d). J2.1-12.1 shows additionally an ubiquitous expression in the whole wing disc (Figure 5.4.b/c).

In the eye disc of all the tested lines the activity is mainly restricted to rows of cells anterior and posterior to morphogenetic furrow (MF) but nearly absent in the furrow (Figure 2.5.3.b/g). In the antennal disc the expression is restricted in the central fold of the disc, the presumptive arista (Figure 2.5.3.b/g).

The leg discs in all lines (Figure 2.5.3.b/j) express *lacZ* in a concentric ring around the center of the disc with a hole in the middle. In addition there are single cells that synthesize lacZ at high level; these cells could be sensory organ progenitor cells (SOP) (Figure 2.5.3.b/j, tip of yellow arrow and Figure 5.4/h). The staining in the leg discs of all five lines is qualitatively and quantitatively similar.

**J3**: four independent lines, namely J3-68E-34.1, J3-68E-36, J3-68E-46.1, and J3-68E-48 were analyzed.

**J4**: four independent transgenic stocks, J4-34.4, J4-101.6, J4-86F-17.1, and J4-86F-22.1 were examined.

Both reporters show an activity similar to J2.1 in the brain of third instar larvae for all the analyzed lines, except that the ubiquitous expression is slightly stronger compared to the J2.1 transgene (Figure 2.5.3.b/b, c).

β-gal activity in the wing imaginal discs of the lines J3-34.1 and J4-34.4 (Figure 2.5.3.b/e, f) is qualitatively similar to the pattern observed for J2.1, except that the J4-34.4 strain shows a slightly ubiquitous expression all over the wing disc (Figure 2.5.3.b/f). In the two examined lines, J3-36 and J3-48, the expression in the notum region is weak (Appendix, Figure 5.4.b/k, l).

In the eye-antennal and the leg discs, *lacZ* activity is qualitatively similar to J2.1, but it is slightly weaker (Figure 2.5.3.b/h, i, k, l). The line J4-86F-17.1 lacks *lacZ* activity in the antennal disc (Appendix, Figure 5.4.b/u).

X-gal staining was carried out in embryos of transgenic fly lines for each reporter construct J2.1, J3, and J4. The embryos collected from J2.1-30.1, J3-68E-46.1, and J4-34.4 parents were stained over-night at room temperature for *lacZ* activity. In the earliest stages all the embryos were stained strongly (black) over-night. For lacZ detection in the later stages, embryos were stained over-night. For the early stages, the experiment was repeated with the embryos recollected from the above flies and stained for three hours at room temperature.

The intensity of *lacZ* activity was observed at the same high level in all the embryos throughout embryogenesis (Figure 2.5.3.b/m-x). In the earliest stages *lacZ* activity can be detected ubiquitously. During germband extension the lacZ mainly accumulates in the midgut and mesoderm. For the staining in positive and negative controls that were conducted in parallel to each staining, see Appendix section 5.2.

For lacZ detection in ovaries for each construct the same independent transgenic lines were analyzed as for the embryos, namely J2.1-30.1, J3-68E-46.1, and J4-34.4. Constructs J2.1 to J4 all show strong *lacZ* (Figure 2.5.3.b/y-z') activity in the germarium and in the nuclei of two cell types in egg chambers, nurse cells and oocytes. However, there is no reporter activity in the follicle cells.

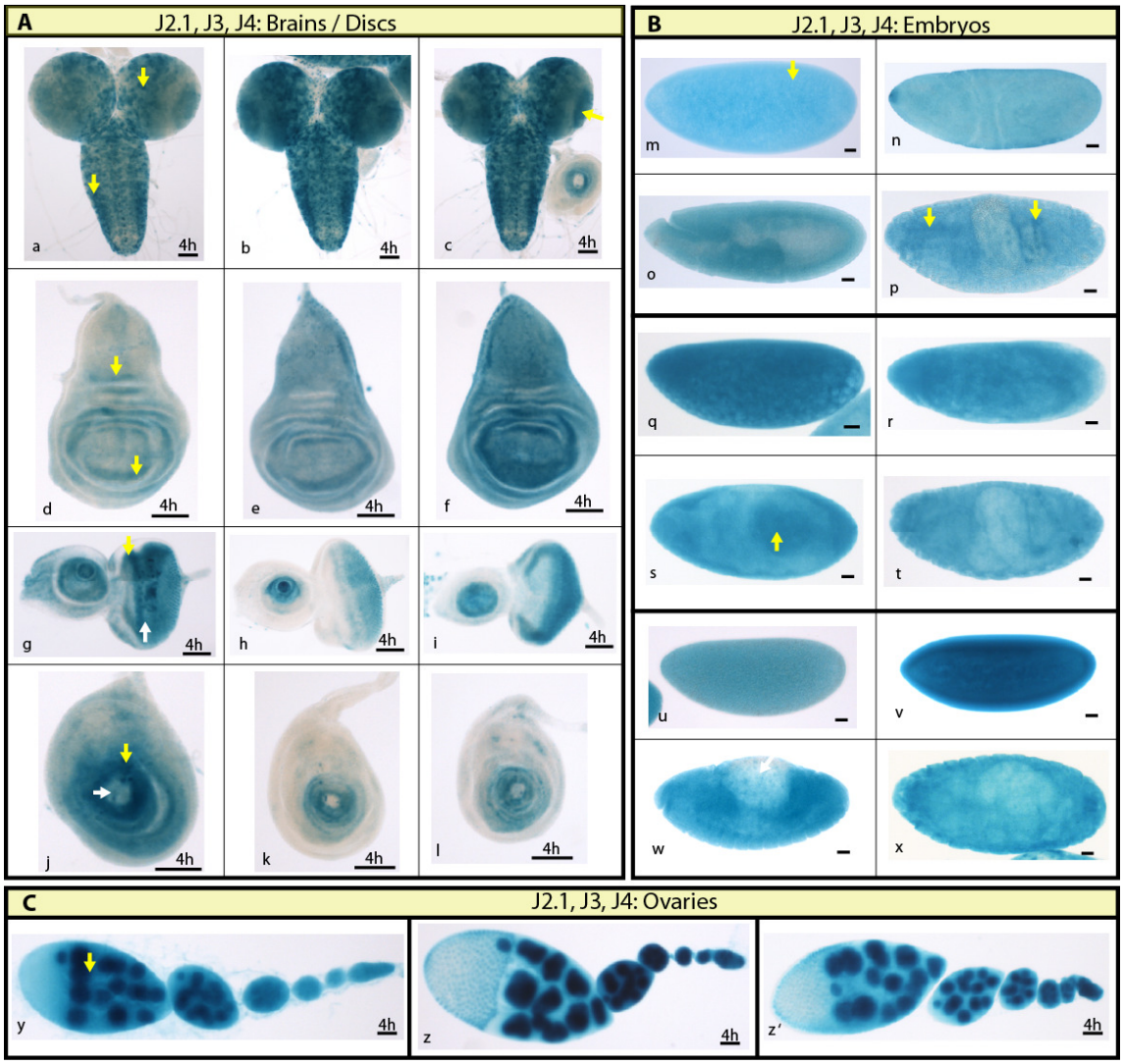



**Figure 2.5.3.b. *LacZ* activity under control of J2.1-J4 deletions from the *dmyc* 5' regulatory region.** (**A**) J2.1 (J2.1-30; **a**, **d**, **g**, **j**) and its deletions J3 (J3-68E-34.1: **b**, **e**, **h**, **k**), and deletion up to 2.7-kb in the J4 construct (J4-34.4: **c**, **f**, **i**, **l**) were able to express the reporter in a *dmyc* manner in the brain and discs (**a**-**c** brain; **d**-**f** wing discs; **g**-**i** eye discs; **j**-**l** leg discs). (**B**) Embryos taken from transgenic animals carrying transgenes either J2.1-30.1 (**m**-**p**), or J3-68E-46.1 (**q**-**t**), or J4-34.4 (**u**-**x**) retained the *dmyc* like expression during different stages (embryo stages: **m**, **n**, **q**, **r**, **u**, **v**: 2–6; **o**, **s**, **w**: 11–12; **p**, **t**, **x**: 13–16). (**C**) In the ovaries, the *dmyc* like expression remained unchanged in all three lines, J2.1-30.1 (**y**), J3-68E-46.1 (**z**), and J4-34.4 (**z'**). Yellow arrow indicates *lacZ* expression and white arrow indicates lack of *lacZ* activity. Staining time for discs and ovary is indicated. Embryos were stained for 25 minutes for detection in early stages (**m**, **n**, **q**, **r**, **u**, **v**), and over-night for later stages. Scale bar in (**a**-**z'**) indicates 50 µm.

**J5**: for this construct the four independent transformants, J5-22, J5-50.1, J5-92, and J5-86F-15.1 were looked at. β-gal activity in the brain is strong along proximal parts of the hemispheres and all over the ventral ganglion (Figure 2.5.3.c/a). Expression of

*lacZ* is qualitatively and quantitatively very similar in the tested lines and is very similar to J2.1 (see also Appendix, Figure 5.4.c/a, b).

In the wing imaginal discs β-galactosidase activity is predominantly around the wing pouch and in the notum region (Figure 2.5.3.b/d). However, like J4 construct (Figure 2.5.3.b/f) the line J5-86F-15.1 shows additionally ubiquitous expression all over the wing disc (Figure 2.5.3.c/d).

Reporter activity in the eye disc (Figure 2.5.3.c/g) is very similar to J2.1, anterior and posterior to MF but very weak in the furrow. The line J5-22 shows a ubiquitous lacZ staining all over the disc (Figure 5.4.c/e).

In the leg disc (Figure 2.5.3.c/j) the staining is evident in concentric ring around the middle of the disc. Additionally, the wing disc in the line J5-86F-15.1 shows ubiquitous staining all over the leg disc.

**J6**: five independent lines, J6-18.2, J6-34.2, J6-58.2, J6-64, and J6-112 were tested. The reporter activity in the brain (Figure 2.5.3.c/b) of J6-18.2 larvae is very strong at the distal part of the both hemispheres. These hemispheric parts could correspond to so-called highly proliferative mushroom body neuroblasts (MBNbs). The tested lines J6-34.2 and J6-112 show strong activity for the *lacZ* reporter in the proximal halves of the hemispheres (Figure 5.4.c/ i, j).

All the tested lines express β-gal in the wing, eye, and leg discs (Figure 2.5.3.c/e, h, k) qualitatively similar to J2.1 (Figure 2.5.3.b/d, g, j), indicating that the J6 truncation still retains the activity in the brain and imaginal tissues. For more staining in the J6 transgenic lines see Appendix, Figure 5.4.c.

**J7**: for this construct the four independent lines, J7-79.1, J7-114, J7-145.2, and J7-6.2 were looked at. None of the examined lines show any β-gal activity, neither in the brain nor in any imaginal discs (eye-antennal, wing, leg) (Figure 2.5.3.c/c, f, i, l). Reverse transcription polymerase chain reaction on the discs shown in Figure 2.5.3.c (I, l), did not detect any *lacZ* transcripts beyond background level in these tissues that were negative for lacZ staining (Figure 2.5.3.cF, lanes –RT to l).

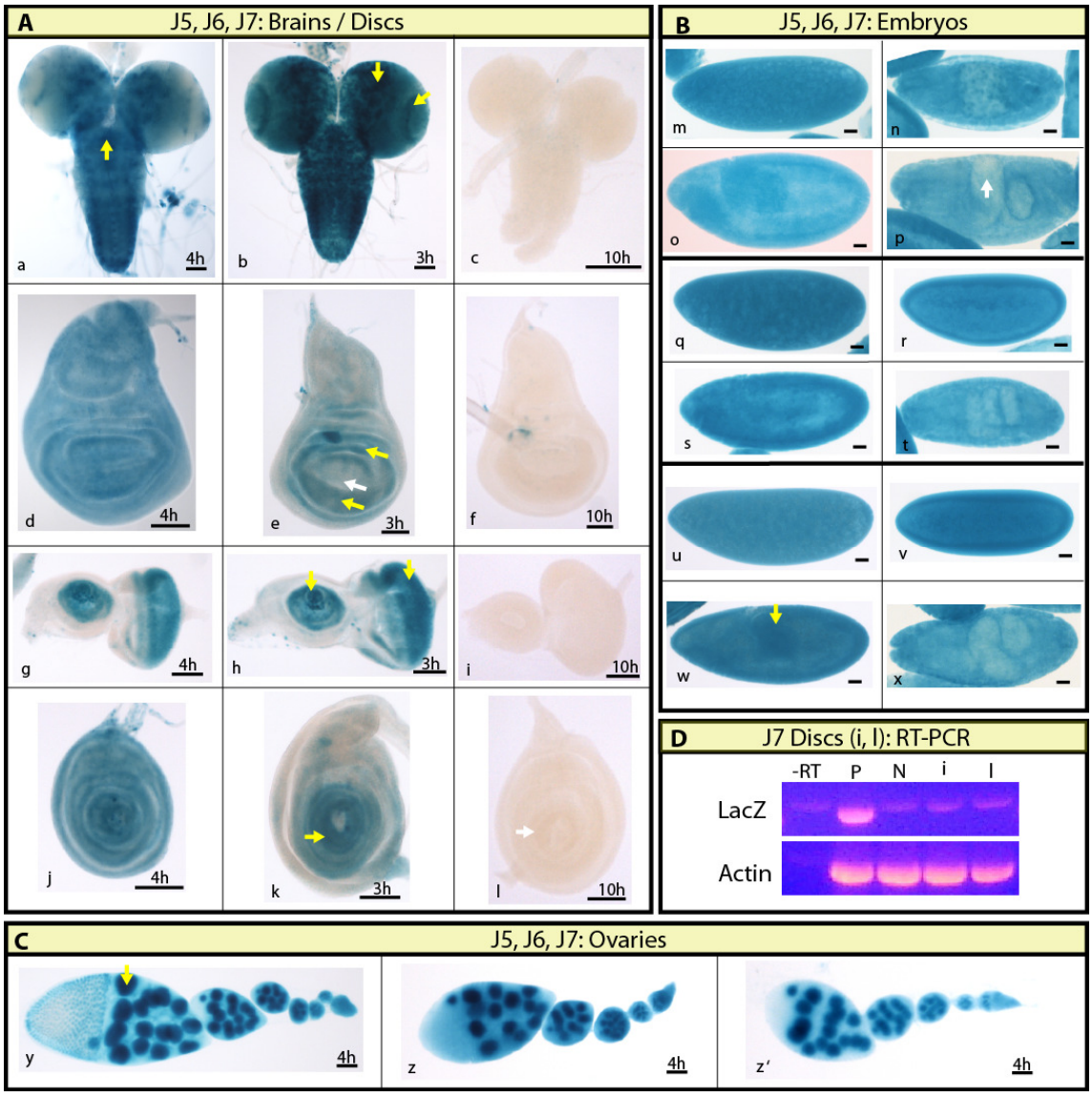



**Figure 2.5.3.c. *LacZ* activity under control of J5-J7 transgenes carrying different deletions from the *dmyc* 5' regulatory region.** (**A**) J5 (J5-86F-15.1: **a**, **d**, **g**, **j**) and the truncation J6 (J6-18.2: **b**, **e**, **h**, **k**) were able to express the reporter in a *dmyc* manner in the brain and discs (**a**, **b** brain; **d**, **e** wing discs; **g**, **h** eye discs; **j**, **k** leg discs). Further truncation down to 1.923-kb of J7 resulted in a loss of expression in the brain and discs (J7-79.1: **c**, **f**, **i**, **l**). (**B**). All tested lines retained the expression in the embryos (J5-86F-15.1: **m**–**p**, J6-58.2: **q**–**t**, J7-79.1: **u**-**x** embryo stages: **m**, **q**, **r**, **u**, **v**: 2–6; **n**, **s**, **w**: 11–12; **p**, **t**, **x**: 13–16). (**C**) Ovaries were taken from the same transgenic animals as in (B). All the constructs recapitulated the *dmyc*-like expression in ovary (**y**-**z'**). Yellow arrow indicates l*acZ* expression and white arrow indicates lack of *lacZ* activity. (**D**) Reverse transcriptase polymerase chain reaction on J7 discs (**i**, **l**, shown in **A**), detects no *lacZ* transcripts beyond background level. Controls used are as follows: -RT, negative control with no transcriptase (brains from **b**, shown in **A**); P, **A**, **b**, positive control; N, *y[1] w[1118]* leg discs, negative control; actin, *Drosophila* actin as internal control. Staining time for discs and ovary is indicated. Embryos were stained for 25 minutes for detection in early stages (**m**, **q**, **r**, **u**, **v**), and over-night for later stages. Scale bar in (**a**-**z'**) indicates 50 µm.

For lacZ detection in the embryos and ovaries for each construct one independent transgenic line was analyzed, namely J5-86F-15.1, J6-58.2, and J7-79.1. The expression in embryos and ovaries is very similar to the staining in the embryos ovaries taken from the J2.1-J4 transgenic animals (Figure 2.5.3.c/y-z').

### 2.5.4 *lacZ* Reporter Activity under Control of the *dmyc* Large Intron

In most protein-coding genes of *Drosophila*, there is a downstream core promoter element that functions cooperatively with an initiator to facilitate the binding of transcription factors in the absence of a TATA box (Kutach and Kadonaga, 2000). The high throughput expression data for the *D. melanogaster* transcriptome, generated by tilling arrays, has shown that most of introns are transcriptionally active within the early hours of development at specific time points (Daines et al., 2011). Scanning of the intron 2 region with *cis*-Decoder, initially revealed clusters of multiple conserved sequence blocks, common to most *Drosophila* species. A search using the Lasergene 9.1 GeneQuest module identified a downstream promoter region with no TATA box, comparable with the *Drosophila* consensus sequence (Figure 2.1.1.b, and see Discussion, section 3.1.4). In order to investigate reporter activity under control of *dmyc* intragenic region, different full-length and deletion constructs were generated from this region (Figure 2.4.3.a and Figure 2.5.4.a).

To study reporter activity, for each construct from the region independent transgenic lines were established.

**J8**: for this construct the eight independent transformants, J8-20, J8-32.4, J8-78.1, J8-8.1, J8-34.1, J8-86F-17.1, J8-86F-2.1, and J8-2A-31 were looked at. **J8.2**: seven independent transgenic fly stocks, J8.2-14, J8.2-17.1, J8.2-39, J8.2-40.1, J8.2-59.1, J8.2-70.1, and J8.2-112 were examined.

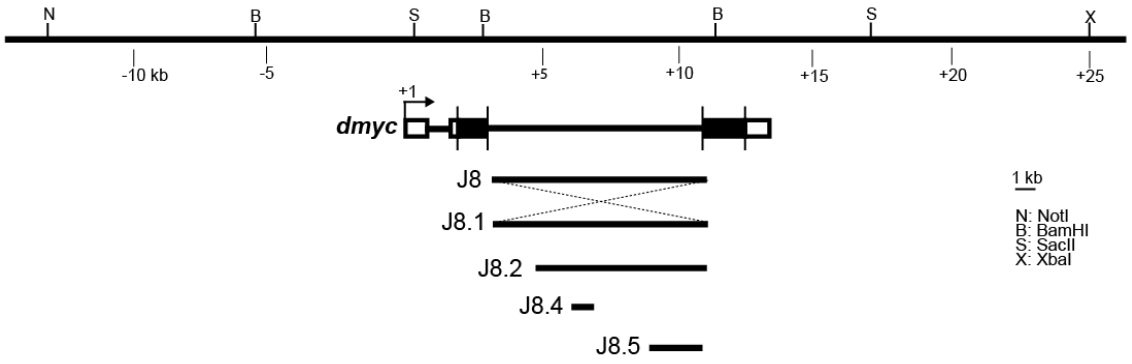



**Figure 2.5.4.a. Deletion constructs generated from the *dmyc* 5' regulatory region are shown.** The 40-kb genomic locus and the location of the *dmyc* are shown at the top. A fragment of ~8.1-kb in size (J8, lacking open reading frame sequences) from the *dmyc* intron 2 region, the reversed full fragment in J8.1, and the deletion restriction fragments (J8.2, J8.4, J8.5) are shown below.

The J8 and the J8.2 transgenes both showed *dmyc*-like expression in brain and imaginal tissues (Figure 2.5.4.b/a–h). During embryogenesis, both transgenes are active in the early and later stages (Figure 2.5.4.b/i–n). However, only the J8.2 transgenic animals, which lack the 2-kb upstream sequences, showed strong patterning of *dmyc* in mesodermal tissues ((Figure 2.5.4.b/o, p). This suggests that the large intron contains regulatory elements that are sufficient for activating the reporter in *dmyc* manner. It further suggests that the upstream sequences might contain elements that influence *dmyc* expression in presumptive mesodermal tissues. Unlike the difference observed in embryonic activity, the expression of *LacZ* in ovarian tissues remains unchanged for both transgenes (Figure 2.5.4.b/q, r).

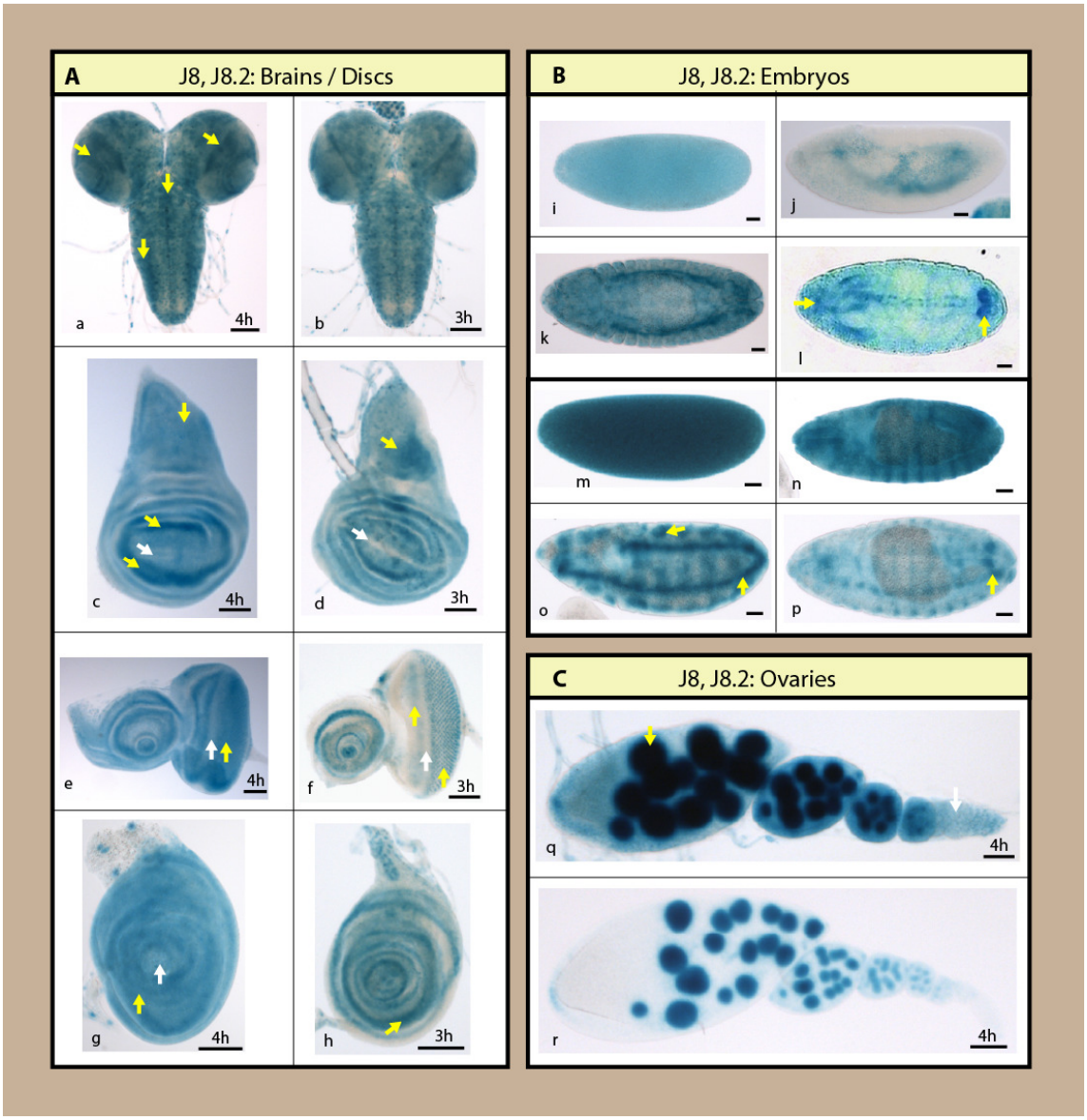



**Figure 2.5.4.b. Analysis of the *dmyc* full-length intron 2 in J8 construct and its 6-kb deletion fragment in J8.2 transgene.** (**A**) The full fragment J8 (J8-20: **a**, **c**, **e**, **g**) and its derived subfragment J8.2 (J8.2-14: **b**, **d**, **f**, **h**) were assayed in larval brain and different imaginal tissues. Both transgenes showed *dmyc*-like expression in these tissues (**a**, **b** brain; **c**, **d** wing discs; **e**, **f** eye discs; **g**, **h** leg discs) (see also raw data, Figure 5.4.d). (**B**) The transgenes J8 and J8.2 both express *lacZ* in early and late embryogenesis (**i**–**n** embryo stages: **i**, **j**, **m**: 2–6; **k**, **l**, **n**: 9–13). However, only the shorter transgene J8.2 shows expression in mesodermal tissues (**o**, **p** embryo stages: **o**: 9-12; **p**: 12–15). (**C**) In the ovary (J8-20, **q**; J8.2-14, **r**) both transgenes are active. Yellow arrow indicates *lacZ* expression and white arrow indicates lack of *lacZ* activity. Staining time for discs and ovary is indicated. Embryos were stained for 25 minutes for detection in early stages (**i**, **j**, **m**), and over-night for later stages. Scale bar in (**a**–**r**) indicates 50 µm.

In order to test the direction of the regulation from this region experimentally, the reversed full-length intron 2 fragment was fused proximal to distal (3' → 5') to the reporter gene (construct J8.1 in Figure 2.4.3.a and Figure 2.5.4.a).

**J8.1**: six independent lines, J8.1-40.3, J8.1-69.4, J8.1-71, J8.1-79.2, J8.1-83.3, and J8.1-127.2 were examined. Analyses of the reporter activity under control of reversed intronic sequences showed the abolition of reporter activity in virtually all tested tissues (Figure 2.5.4.c/a–i). This observation suggests that the core promoter element only functions in the proximal to distal direction.

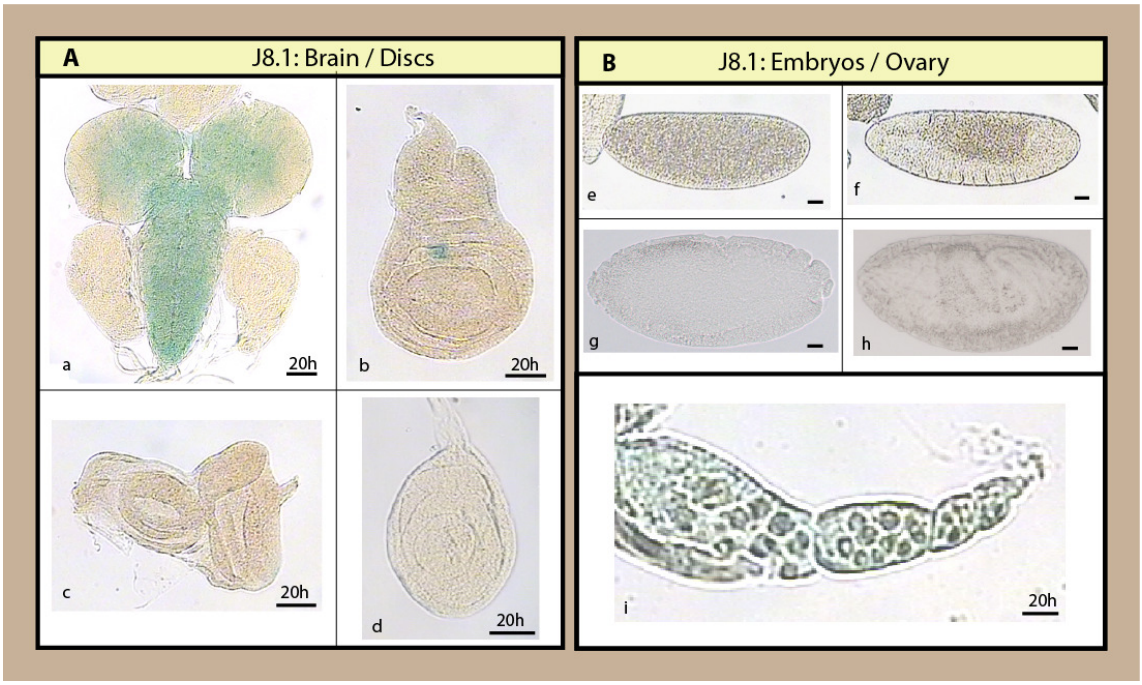



**Figure 2.5.4.c. The 8-kb intron 2 full-length in J8.1 construct functions unidirectionally.** (**A**) The J8.1 fragment (J8.1-71) was examined for its ability to drive reporter expression in different tissues during early developmental stages. Except for a minimal basal expression in the brain (**a**) the activity of the reporter was abolished in the wing, eye, and leg imaginal discs **(b-d)**. (**B**) No expression was detected in the embryos (**e**: stage 1-2; **f**, **g**: stage 9-12; **h**: stage 13-16) and ovary (**i**), compared to the J8 activity in Figure 2.5.4.b. Staining times for the discs and ovary are indicated above the scale bars. Embryo staining took place over-night. Scale bar in (**a-i**) indicates 50 µm.

To gain more insight into the *dmyc* regulation from this downstream regulatory region, the J8.2 truncation was dissected into its cluster of binding sites region and

the sequences downstream containing a putative DPE element, (J8.4 and J8.5 transgenes in Figure 2.4.3.a and Figure 2.5.4.a).

**J8.4**: for this construct the seven independent transformants, J8.4-86F-1, J8.4-86F-5, J8.4-8622, J8.4-25709-3, J8.4-25709-7, J8.4-25710-3, and J8.4-25710-5 were looked at. **J8.5**: six independent lines, J8.5-86F-1, J8.5-86F-4, J8.5-25709-1, J8.5-25709-3, J8.5-25710-1, and J8.5-25710-5 were analyzed.

In β-gal assays for *lacZ* expression under control of each of these elements no activity was detectable beyond basal levels in larval brain and imaginal tissues (Figure 2.5.4.d/a-h). Similarly, the staining in the embryos (Figure 2.5.4.d/i-p) did not exceed the basal level. And in adult *Drosophila* ovaries (Figure 2.5.4.d/C q, r) β-gal reaction was clearly negative. Reverse transcription polymerase chain reaction (Figure 2.5.4.d/D) on the J8.4-86F-5 brain and ovary (Figure 2.5.4.d/a, q shown in A, C) and J8.5-25709-3 disc (Figure 2.5.4.d/d, shown in A) did not detect any *lacZ* transcripts beyond background level in these tissues that were negative for lacZ staining.

To further confirm the obtained results, protein extracts were isolated from the ovarian tissues of transgenic animals carrying either the largest construct J8 or the 3'-end deletion, J8.5 transgene (Figure 2.5.4.e/A, B), and used for the detection of lacZ protein. As expected, the immunoblotting assay was positive for the full-length J8 fragment (Figure 2.5.4.e/A b, lane 2) but turned out to be absent in the J8.5 extracts (Figure 2.5.4.e/A b, lane 3). This result is consistent with the β-gal assay for the detection of lacZ protein in the ovaries (Figure 2.5.4.e/C g, h), positive in the ovary taken from the J8 transgenic flies and negative in the flies carrying the J8.5 construct. These results suggest that the upstream regulatory binding sites in J8.4 transgene are necessary for the activity of downstream regulatory elements in J8.5 fragment for the patterning of *dmyc* from this region.

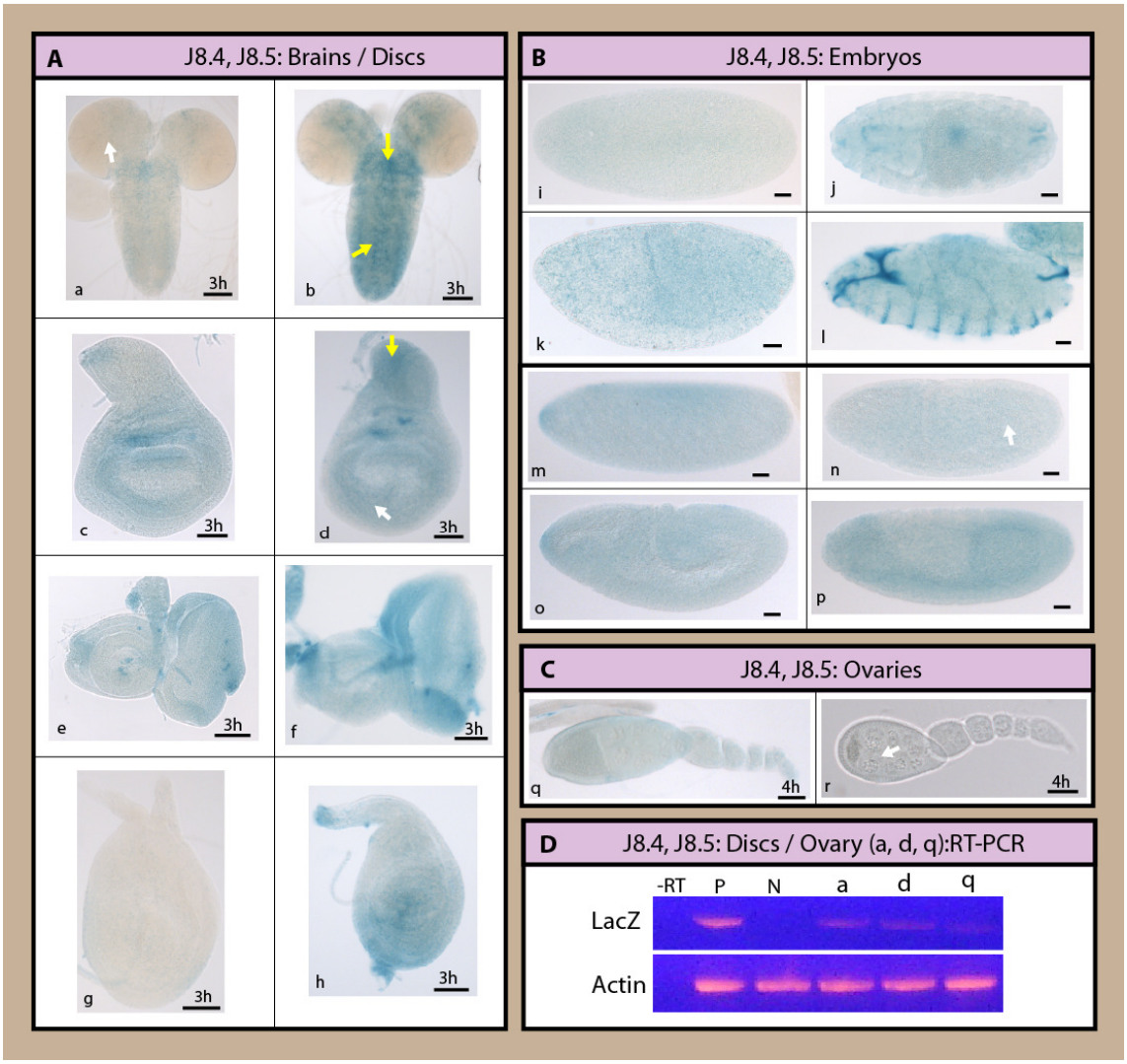



**Figure 2.5.4.d. Reporter activity under control of intronic deletions J8.4 and J8.5.** (**A**) The J8.4 (J8.4-86F-5: **a**, **c**, **e**, **g**) and the large intron 3'-end truncation J8.5 (J8.5-25709-3: **b**, **d**, **f**, **h**) were unable to express the reporter beyond the background level in brain and discs (**a**, **b** brain; **c**, **d** wing discs; **e**, **f** eye discs; **g**, **h** leg discs). (**B**) Both transgenes do not express *lacZ* in early and late embryos (J8.4-86F-5: **i**-**l**; J8.5-25709-3: **m**-**p** embryo stages: **i**, **m**, **n**: 2–6; **j**, **k**, **o**, **p**: 9–12; **l**: 13-16). (**C**) Also in the ovaries the expression remained negative for both constructs (J8.4-86F-5: **g**; J8.5-25709-3: **h**). (**D**) Reverse transcriptase polymerase chain reaction on J8.4-86F-5 brain and ovary (**a**, **q** shown in **A**, **C**) and J8.5-25709-3 disc (**d** shown in **A**), detects no *lacZ* transcripts beyond background level. Controls used are as follows: -RT, negative control with no transcriptase (brain from Figure **2.5.4.b**/**b**); P, **2.5.4.b**/**b**, positive control; N, *y[1] w[1118]* leg discs, negative control; actin, *Drosophila* actin as internal control. Yellow arrow indicates low level of *lacZ* expression and white arrow indicates lack of *lacZ* activity. Staining time for discs and ovary is indicated above the scale bar, and embryos were stained over-night. Scale bar in (**a**–**r**) indicates 50 µm.

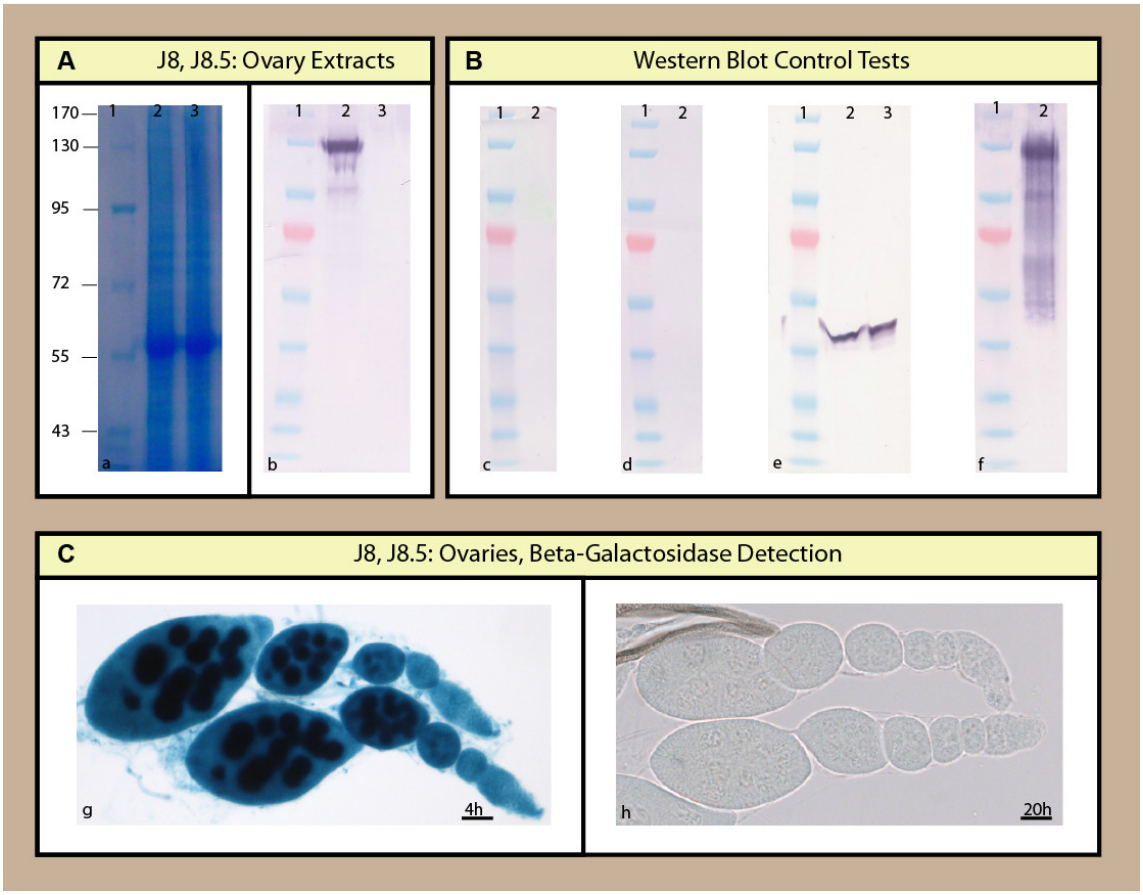



**Figure 2.5.4.e. Western Blot analysis of the ovary extracts isolated from flies carrying J8 or J8.5 transgenes.** (**A**) Proteins extracted from the ovaries of the female flies carrying J8 (J8-32.4) or J8.5 (J8.5-25710-1) transgenes were analyzed by SDS-PAGE and immunoblotting. (**a**) Gel SafeBlue staining of the protein extracts: **1**: protein marker in kDa unit; **2**: J8 lysate; **3**: J8.5 lysate. (**b**) The extracts in (**a**) were subjected to immunoblotting with antibody anti-beta gal. lacZ protein was detectable in the J8 ovaries (panel **b**, lane **2**), whereas, J8.5 extract was negative (panel **b**, lane **3**). (**B**) Following control blots were performed: (**c**) J8 extracts no primary antibody added, (**d**) J8 extracts no secondary antibody added, (**e**) extracts in (**a**) were subjected to immunoblotting with antibody against *Drosophila* actin, (**f**) positive Control: beta-gal protein. (**C**) lacZ staining of ovaries taken from J8-32.4 (**g**) and J8.5-25710-1 (**h**). The staining time for ovaries is indicated above the scale bar. Scale bar in (**g**, **h**) is 50 µm.

### 2.5.5 Fusion of the Upstream Promoter Region to the *dmyc* Large Intron

In the chimeric construct J9 the fully active J5 deletion is joined in frame to the full-length intron 2 fragment, whereas in J10 construct the J5 promoter is fused to the intron 2 sequence in reverse orientation (Figure 2.5.5.a).

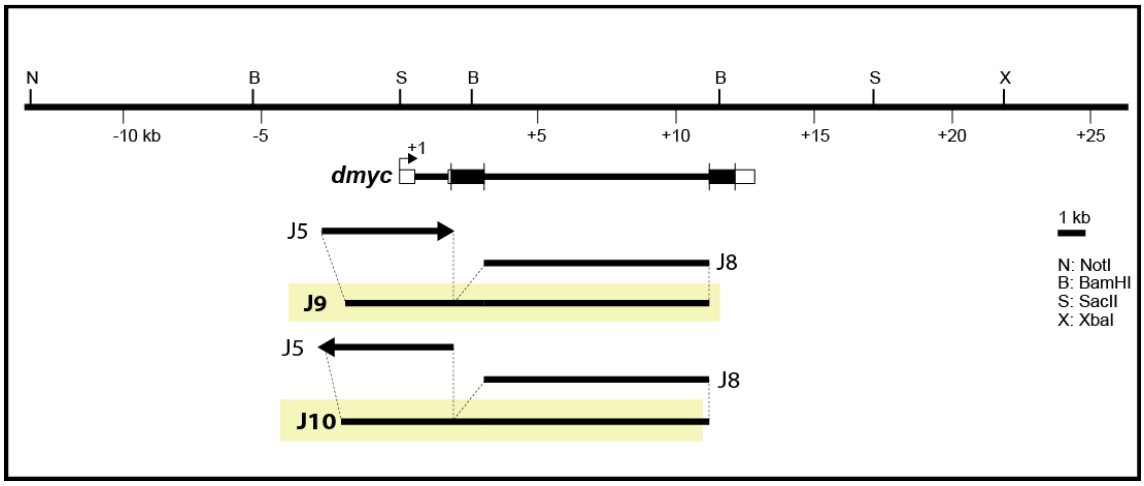



**Figure 2.5.5.a. Schematic depiction of J9 and J10 constructs.** The 40-kb genomic locus and the location of the *dmyc* gene are shown at the top. The fragments J5 and J8 (both lacking open reading frame sequences) are shown below. J5 is fused 5' → 3' to the intron 2 in J9 and 3' → 5' in J10 transgene (see also Figure 2.4.3.b).

**J9**: three independent lines, namely J9-79.1, J9-89.5, and J9-108.1, were analyzed. **J10**: the five independent transgenic stocks, J10-34.1, J10-26.1, J10-57.1, J10-60, and J10-93.1 were examined. All three lines containing J9 construct express the reporter in the brain and discs similar to J2.1-J6 and J8, J8.2 constructs (Figure 2.5.5.b/a, c, e, g). The fly lines J10-34.1, J10-26.1, J10-57.1, and J10-93.1 express the reporter in the larval brain and discs similar to J9 flies (Figure 2.5.5.b/b, f, h), except the line J10-60 (Figure 2.5.5.b/d) that shows a weaker activity in the wing disc compared to J9 and the other four lines in J10.

However, analysis of embryos and ovaries taken from the J9-79.1, J10-34.1, and J10-60 flies, showed partial attenuation of *dmyc* embryonic activity in the J10 fly lines (Figure 2.5.5.b/i-p). The mode of action of enhancer elements responsible for embryonic development was dependent on the spacing and arrangement of binding sites, with the activity in imaginal tissues and ovaries remaining unaffected.

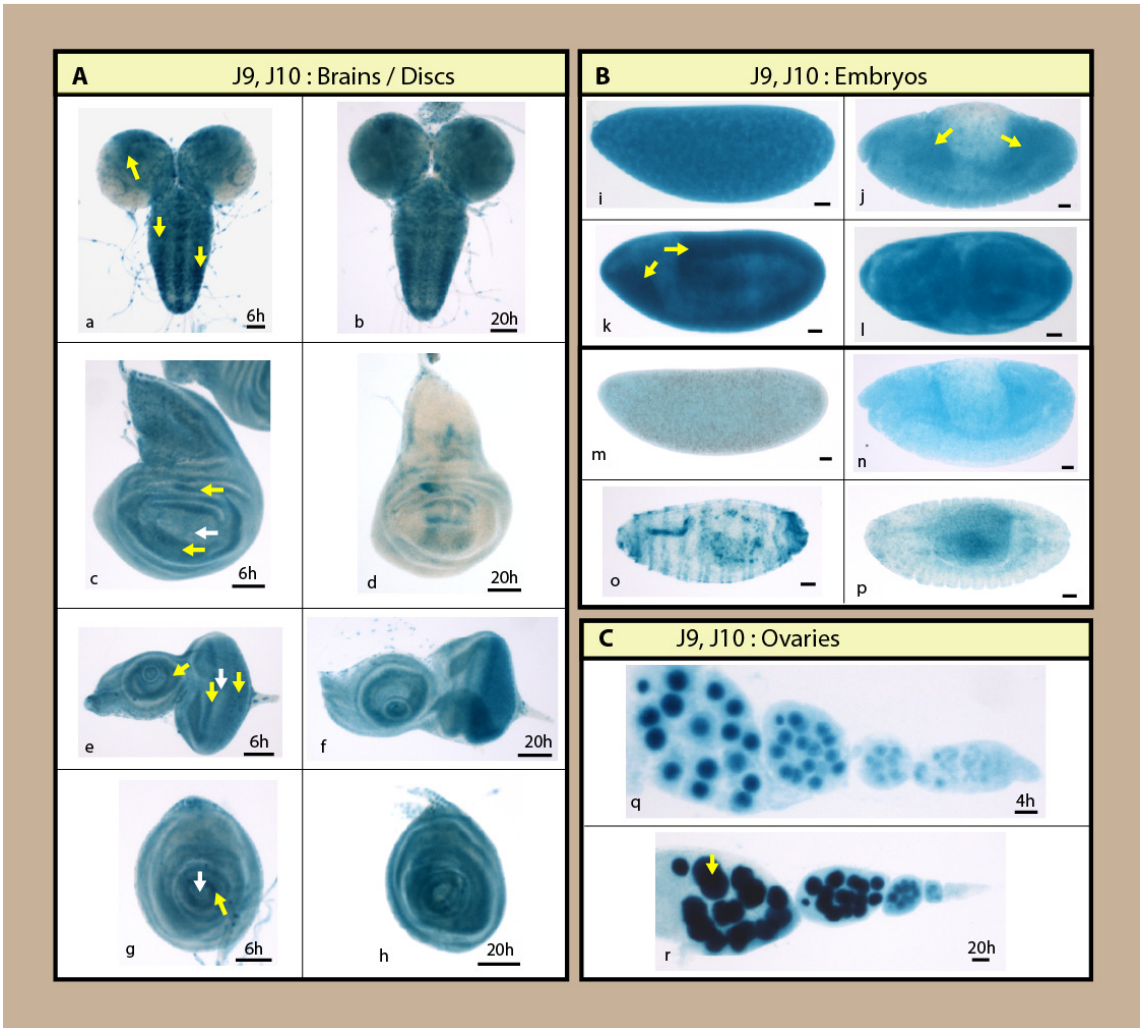



**Figure 2.5.5.b. *lacZ* activity under control of transgenes J9 and J10.** (**A**) Brain and imaginal discs in the J9 and J10 transgenes are taken from the fly lines J9-79.1 (**a**, **c**, **e**, **g**), J10-34.1 (**b**, **f**, **h**), and J10-60 (**d**). Reporter activity in the larval tissues taken from the J9-79.1 and J10-34.1 remained unchanged (**a**, **b** brain; **c**, wing disc; **e**, **f** eye discs; **g**, **h** leg discs), except the line J10-60 that was weaker in the wing disc (**d**). (**B**) In the embryos J9 expression did not change (**i**–**l**), however, *lacZ* activity in J10 was weaker (**m**–**p**) (embryo stages: **l**, **m**: 1–4; **j**, **k**, **l**, **p**: 9–12; **p**: 13–15). (**C**) In ovaries (J9-79.1, **q**; J10-34.1, **r**), the expression remained unchanged and similar to J2.1 and its deletions, J8, and J8.2 transgenes (Figure 2.5.3.b, 2.5.3.c, and 2.5.4.b). Yellow arrow indicates *lacZ* expression and white arrow indicates lack of *lacZ* activity. Staining times for discs and ovaries are indicated above the scale bar. Embryo shown in (**i**) was stained for 25 minutes, and the rest were stained over-night. Scale bar in (**a**–**r**) indicates 50 µm.

### 2.5.6 Analysis of the *dmyc* 3'-End Fragments

In addition to understanding transcriptional activation of *dmyc* and deciphering the nature of the involved *cis*-elements, I was interested in understanding the mechanism of transcript termination. A search with the bioinformatics tool DNASTAR Lasergene

GeneQuest module, detected three potential polyadenylation signals in this region (see Discussion, section 3.1.5). On the basis of identified polyadenylation signals deletion fragments were generated from the *dmyc* 3' region. After injection of these truncations (Figure 2.4.2.d and Figure 2.5.6.a.), for each construct independent transgenic fly stocks were established and analyzed.

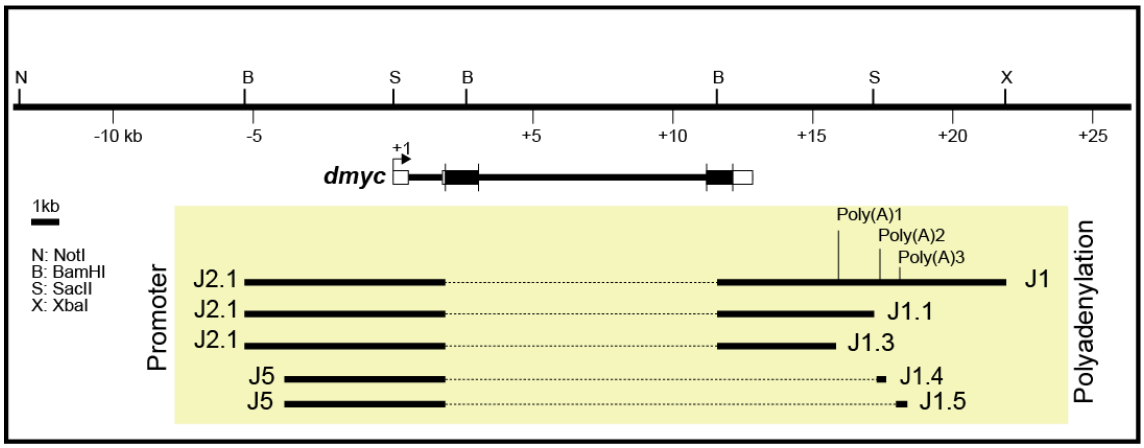



**Figure 2.5.6.a. Depiction of the fragments from the *dmyc* 5' and 3'-ends in the reporter constructs J1, J1.1, and J1.3-J1.5.** Relative location of the *dmyc* gene to the 40-kb genomic region is shown on top. The full-length fragment from the *dmyc* 3' region in the J1 construct and its deletions are shown below. The transgenes J1, J1.1, and J1.3 are under the control of the full-length 5'-promoter (see also J2.1 in Figure 2.4.2.b). The constructs J1.4 and J1.5 contain the promoter of J5 in Figure 2.4.2.b.

**J1**: for this construct the two independent transformants, J1-92.9 and J1-29 were looked at.

**J1.1**: for this construct the four independent transformants, J1.1-17, J1.1-19.2, J1.1-2.2, and J1.1-29 were tested.

The staining in the brain of third instar larvae of both constructs is restricted to a limited number of cells distributed in the two proximal halves of the hemispheres and along the ventral ganglion (Figure 2.5.6.b/A a, b). There is a faint ubiquitous activity in the larval brain of both analyzed lines, with an increased level of expression in the middle of the front part of both hemispheres and along the proximal part of the ventral ganglion. Considering the number of cells staining for lacZ, the expression pattern observed in the brain of both analyzed transgenic lines, J1 and J1.1, is highly consistent. However, the ring of cells at the so-called mushroom body neuroblasts is more evident in the J1.1 brains than the J1 (Figure 2.5.6.b/A b).

The analyzed lines express the reporter mainly around the wing pouch (Figure 2.5.6.b/A c, d). The expression in notum region is faint in the lines analyzed for both constructs.

In the eye disc of all the tested lines the activity is mainly restricted to rows of cells anterior and posterior to morphogenetic furrow (MF) (Figure 2.5.6.b/A e, f). In the antennal disc the expression is restricted in the central fold of the disc. In the eye-antennal disc the expression under control of J1 and J1.1 transgenes is very similar to J2.1.

In the leg disc of the tested lines for both transgenes a strong reporter activity is detectable in a concentric ring around the center of the disc with a hole in the middle and this expression is similar to the pattern seen with J2.1 (Figure 2.5.6.b/A g, h).

For *lacZ* activity in the embryos and ovaries, the two transgenic lines J1-92.9 and J1.1-19.2 were analyzed. In the early embryonic stages reporter activity can be detected ubiquitously (Figure 2.5.6.c/i, m, n). In later stages, reporter is mainly active in pharynx, and during germband extension in the midgut and hindgut (Figure 2.5.6.b/j-l, n, o, p). In the ovary, both analyzed lines show strong *lacZ* activity in the germarium and in the nuclei of two cell types in egg chambers, nurse cells and oocytes (Figure 2.5.6.b/q, r).

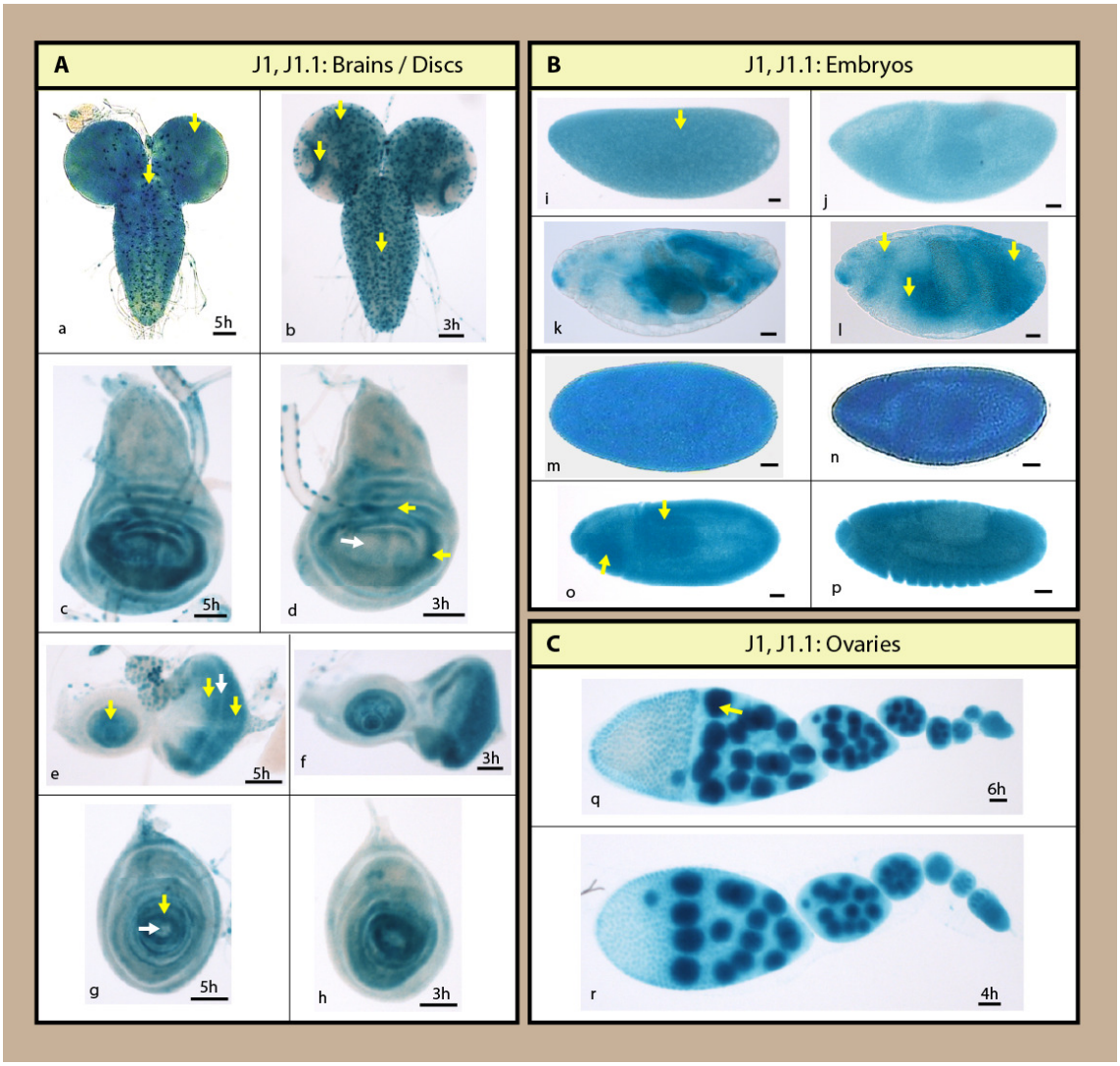



**Figure 2.5.6.b. Reporter activity in the constructs J1 and J1.1.** (**A**) Both constructs (J1-92.9: **a, c, e, g**; J1.1-19.2: **b, d, f, h**) are capable of mediating regulated expression of the reporter in a manner similar to *dmyc* expression in the brain and imaginal tissues, brain (**a**, **b**) and discs (**c**, **d** wing; **e**, **f** eye; **g**, **h** leg). (**B**) Both transgenes are active during different stages of embryogenesis (J1-92.9: **i**-**l**; J1.1-19.2: **m**, **n**; J1.1-29: **o**, **p** embryo stages: **i**, **m** 2–5; **j**, **n**, **o**: 9–11; **k**, **l**, **p**: 12–15) and in (**C**) ovaries (J1-92.9, **q**; J1.1-19.2, **r**). Yellow arrow indicates *lacZ* expression and white arrow indicates lack of *lacZ* activity. Staining times for discs and ovaries are indicated above the scale bar, and embryos were stained for 25 minutes for detection in early stages (**i**, **j**, **m**, **n**, **o**), and over-night for later stages. Scale bar in (**a**–**r**) indicates 50 µm.

**J1.3**: five independent transgenic fly lines, J1.3-12.1, J1.3-17.2, J1.3-22.3, J1.3-40.5, and J1.3-62.1 were examined. The staining in the brain of the analyzed line J1.3-12.1, J1.3-22.3, J1.3-40.5, and J1.3-62.1 is intensive in single cells of both hemispheres and along the proximal part of the ventral ganglion (Figure 2.5.6.c/a).

Also the ubiquitous staining in these areas is stronger. In the line J1.3-17.2 staining is very similar to the other lines, except the ubiquitous staining is weaker.

The expression in the wing disc is detected around the wing pouch, and is weak in the notum (Figure 2.5.6.c/d). The pattern observed is very similar to the expression in J1.1.

In the eye-antennal disc *lacZ* is expressed along both sides of the furrow and in the central ring of antennal disc (Figure 2.5.6.c/g).

In the leg disc reporter is active in concentric rings around the center of the disc but absent from the center (Figure 2.5.6.c/j).

**J1.4**: for this construct the five independent transgenic stocks, J1.4-13, J1.4-14.2, J1.4-15.2, J1.4-17, and J1.4-28.2 were investigated. The expression in the brain is restricted to a limited number of cells distributed in the two proximal halves of the hemispheres and along the ventral ganglion (Figure 2.5.6.c/b).

In the wing disc the activity is strong around the wing pouch but faint in the notum region (Figure 2.5.6.c/e). The staining is qualitatively very similar to J1.3; however, it is slightly stronger.

In the eye-antennal disc *lacZ* is synthesized anterior and posterior to MF and in the center of antennal disc (Figure 2.5.6.c/h).

In the leg discs of the tested lines (Figure 2.5.6.c/k) *lacZ* is expressed in a concentric ring around the center of the disc with a hole in the middle.

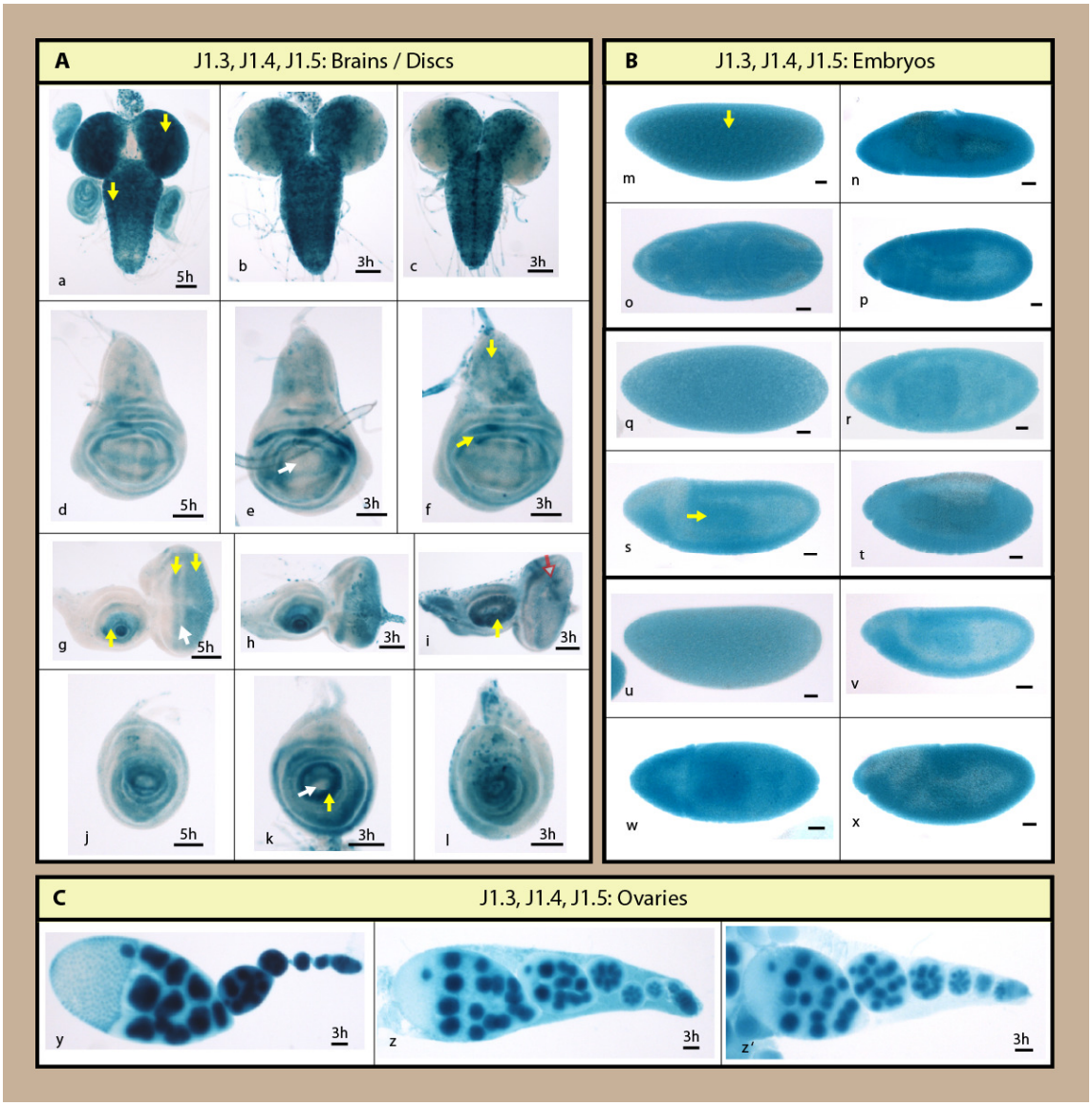



**Figure 2.5.6.c. Reporter activity in the constructs J1.3-J1.5.** (**A**) Brains and discs are taken from the following lines: J1.3-12.1: (**a**, **d**, **g**, **j**); J1.4-14.2 (**b**, **e**, **h**, **k**); J1.5-64.2 (**c**, **f**, **i**, **l**). Shown is the activity of the reporter in the brain and imaginal tissues, brain (**a**, **b**, **c**) and discs (**d**, **e**, **f** wing; **g**, **h**, **i** eye; **j**, **k**, **l** leg). (**B**) The expression of *lacZ* during different stages of embryogenesis is shown (J1.3-12.1: **m**-**p**; J1.4-14.2, **q**-**t**; J1.5-64.2, **u**-**x** embryo stages: **m**, **q**, **u**: 2–5; **n**-**p**, **r**, **s**, **t**, **v**: 9–12; **w**, **x**: 13–15). (**C**) Expression in ovaries is depicted (J1.3-12.1, **y**; J1.4-14.2, **z**; J1.5-64.2, **z'**). Yellow arrow indicates *lacZ* expression and white arrow indicates lack of *lacZ* activity. Unspecific staining of macrophages is indicated with red arrow. Staining times for discs and ovaries are indicated above the scale bar, and embryos were stained for 25 minutes for detection in early stages (**m**, **o**, **q**, **u**), and over-night for later stages. Scale bar in (**a**–**z'**) indicates 50 µm.

**J1.5**: five independent transformants, namely J1.5-64.2, J1.5-75.1, J1.5-81, J1.5-89.3, and J1.5-96 were looked at. The staining in the brain is restricted to a limited

number of cells distributed in the two proximal halves of the hemispheres and along the ventral ganglion, and this activity is similar to J1.4 (Figure 2.5.6.c/c).

In the eye disc of all the tested lines the activity is mainly restricted to rows of cells anterior and posterior to morphogenetic furrow (MF) but nearly absent in the furrow (Figure 2.5.6.c/i). In the antennal disc the expression is restricted in the central fold of the disc, the presumptive arista (Figure 2.5.6.c/i).

In the leg discs of the tested lines (Figure 2.5.6.c/l) *lacZ* is expressed in a concentric ring around the center of the disc with a hole in the middle. In addition there are single cells that synthesize lacZ at high level. The pattern observed is qualitatively similar to J2.1 (Figure 2.5.3.b/j), but slightly stronger than the J2.1.

For the investigation of reporter activity under control of J1.3-J1.5 transgenes in the embryos and ovaries, the transgenic lines J1.3-12.1, J1.4-14.2, and J1.5-64.2 were analyzed. In the early embryos *lacZ* activity can be detected ubiquitously (Figure 2.5.6.c/m, q, u). In later stages, reporter is active in pharynx, mid-, and hindgut (Figure 2.5.6.c/n-p, r-t, v-x). In the ovary, *lacZ* is expressed in the germarium and in the nuclei of two cell types in egg chambers, nurse cells and oocytes (Figure 2.5.6.c/y-z'). The staining observed is similar to the staining done with the ovaries taken from the J1 and J2.1 fly lines.

## 2.6 Investigation of the *dmyc* Transcription Initiation from the 5'-UTR

Bioinformatics analysis of the sequences adjacent to the *dmyc* exon1 in the 5'-UTR predicted three potential initiation sites in this region (see Discussion, section 3.1.3). Therefore, I was interested in determining the transcription initiation site from this region experimentally by examining the *dmyc* cDNA end. Addition of a poly (A) tail to the 3'-end of an RNA molecule or a mixture of RNA molecules can facilitate studies requiring RNA. One of the positive impacts is to increase the stability of the RNA and enhance its ability for the synthesis of first-strand cDNA (Drummond et al., 1985; Krug and Berger, 1987). In addition to poly (A)-tailing, it is well known that addition of an oligonucleotide to the 5' end of the mRNA before reverse transcription reaction, ensures generation of a full-length cDNA (Zhu et al., 2001). Consequently, the stable and 5' end extended RNA molecule provides a tool for capturing most of its 5' end

that helps to identify potential *cis*-acting regulatory elements and the existence of one or more transcription initiation site(s) (Zhu et al., 2001).

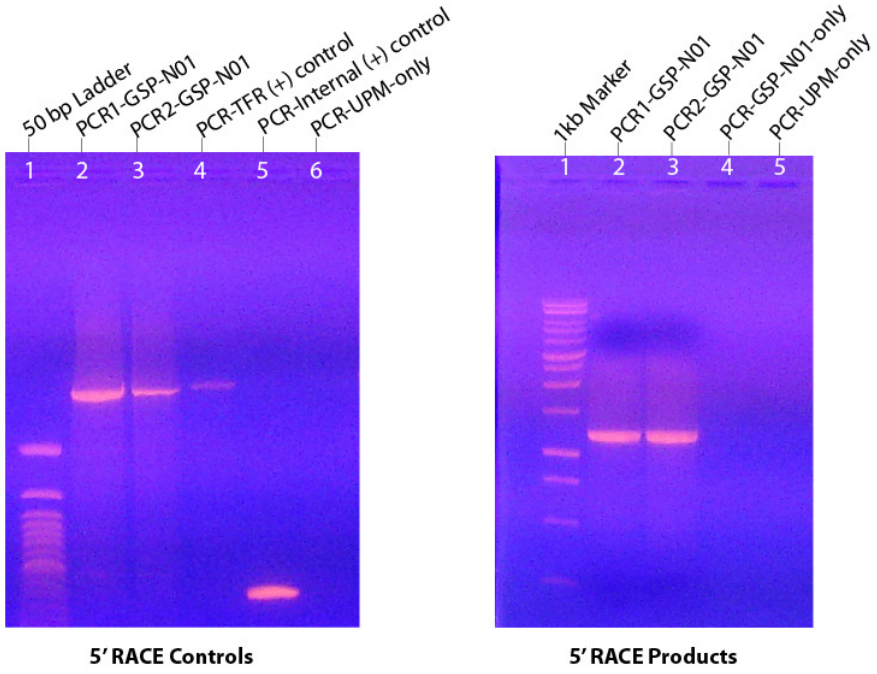



**Figure 2.6.a. Extension of the *dmyc* cDNA at its 5' end.** Polymerase chain reaction amplification of the *dmyc* cDNA 5' end was performed with the forward primer GSP-N01 and the reverse primer UPM, and with two different settings of annealing and extension temperatures (see Materials and Methods, Section 4.6). Sequences for the oligonucleotides are listed in Table 5.1. TFR: Transferrin Receptor; UPM: Universal Primer Mix; GSP: Gene Specific Primer.

To investigate 5' UTR region of the *dmyc* gene, I used ready-to-use *Drosophila* adult poly ($A^+$) mRNA to extend the 5' end of the *dmyc* cDNA (Figure 2.6.a and see Materials and Methods, section 4.6). Analysis of the amplified *dmyc* cDNA revealed that transcription of the *dmyc* gene is most likely to be initiated with an A nucleotide just adjacent to the first initiator element (Figure 2.6.b and Figure 2.6.c).

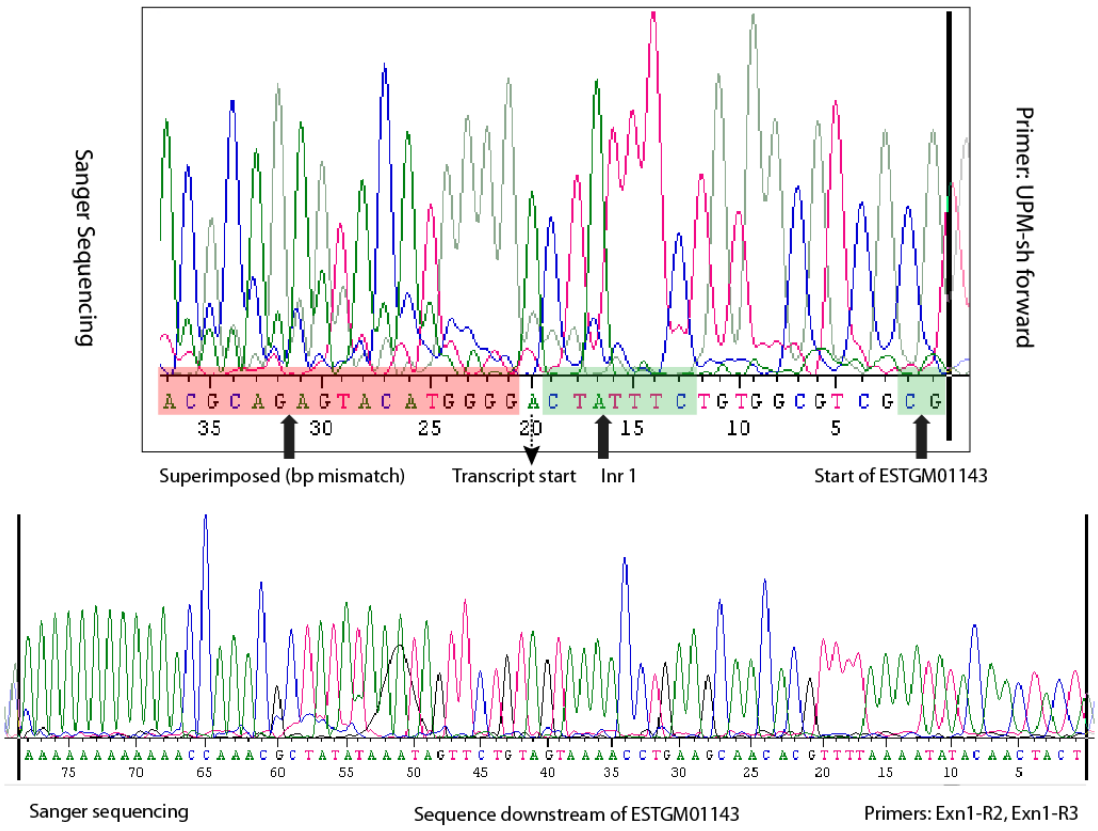



**Figure 2.6.b. Sanger sequencing of the PCR products derived from rapid amplification of the *dmyc* cDNA at its 5' end.** Chromatograms on the top: GSP-N01, the *dmyc* gene specific reverse primer, hybridizes to the *dmyc* exon1 and extends the exon 1 cDNA from its 3'- towards the 5'-end. With this primer two polymerase chain reactions with different settings for annealing and extension temperatures were performed (see Materials and Methods). The PCR products were sequenced with the forward primer UPM-sh, the reverse primers Exn1-R2, and Exn1-R3. Except an A residue upstream of the Inr 1, no further exact base pair matching to the *dmyc* genomic template strand could be sequenced (red highlight on the chromatogram). The extended region with mismatch is due to ligation of an oligonucleotide to the 5' end of RACE cDNA to ensure capturing of the transcript's start site. Chromatograms at the bottom: GSP-C05 and GSP-C06 hybridize to the *dmyc* exon3 and extend the full-length cDNA from its 3'- towards the 5'- end. The PCR products were sequenced with the reverse primer Exn2-R2 and the forward primer Exn2-F1. The PCR products were sequenced at Microsynth, Switzerland. Splice junctions, Inr 1 and EST are highlighted in green on the chromatograms.

The A residue at the beginning of *dmyc* cDNA, 24 nucleotides downstream of TATA box1 (Figure 2.6.c), appears to be a potential candidate that precedes the cap structure at the 5' end of the transcript. The GC1 box (5'-GCGCGGC-3') (SP1 binding site), TATA box1 (5'AAATTTTATTTAA-3'), and the Initiator 1 element (5'-CTATTTCT-3') may serve as basal promoter elements in this region (Figure 2.6.c).

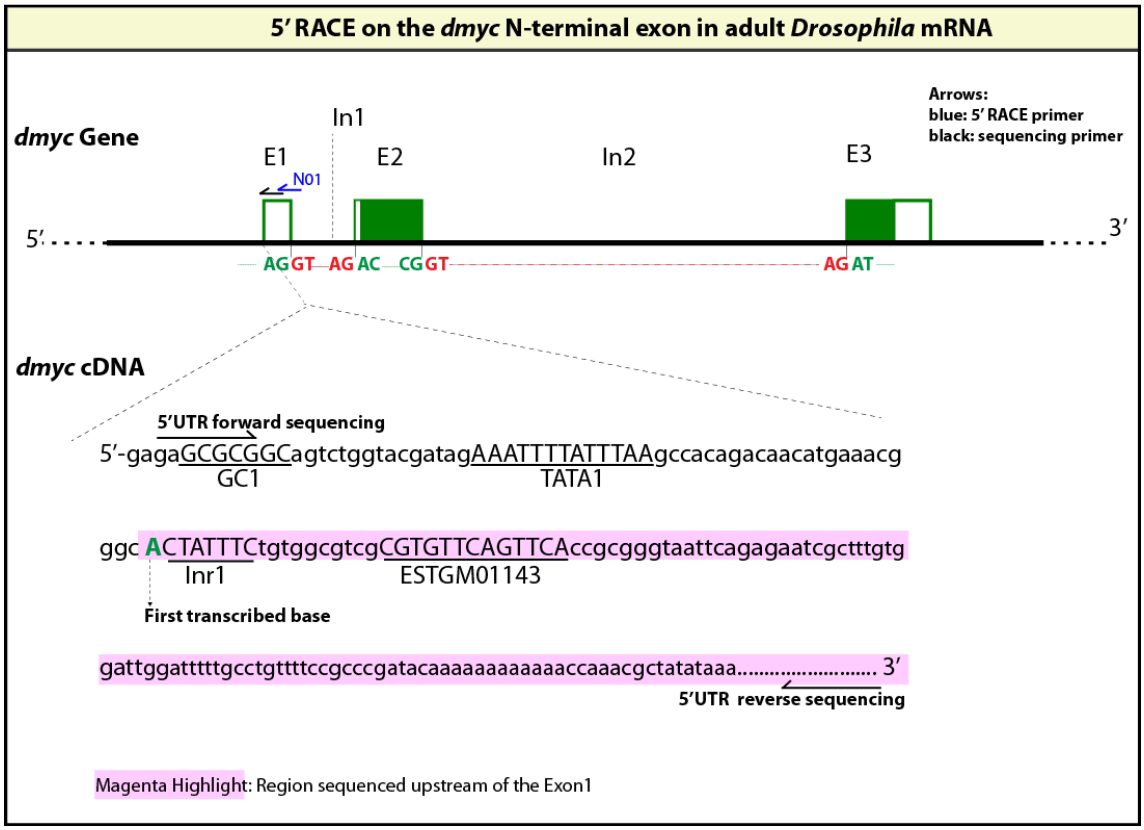



**Figure 2.6.c. 5' RACE analysis of *dmyc* cDNA at its 5' site detects TATA box 1 as main promoter region.** Top: *dmyc* genomic organization (not drawn to scale) with the nucleotides at the exon/intron transitions are shown. Gene specific primer GSP-N01 (blue arrow), used for rapid amplification of the *dmyc* cDNA 5' end, and the sequencing primers (black), used to sequence the *dmyc* cDNA end, are indicated on top of the exon 1 (Note: UPM-sh, the forward sequencing primer is not plotted). Bottom: 5' end of the *dmyc* cDNA. Sequenced portions of the cDNA 5' end are highlighted in magenta and first transcribed nucleotide is in green. For the sequences of primers see Table 5.1 and for the sequencing chromatograms see Figure 2.6.b.

An evidence for the usage of above core promoter elements for the *dmyc* transcription initiation comes from the fact that the insertion of the P-element *P{lacW}* in the enhancer trap line $w^{67c23}$ *P{lacW}*$dm^{G0354}$ (Figure 2.6.d) causes the affected allele to be lethal (McQuilton et al., 2012; Mitchell et al., 2010). Analysis of the insertion breakpoint revealed that the P-element insertion leads to loss of 56 nucleotides in this region, including TATA box1 and GC box1 (Figure 2.6.d).

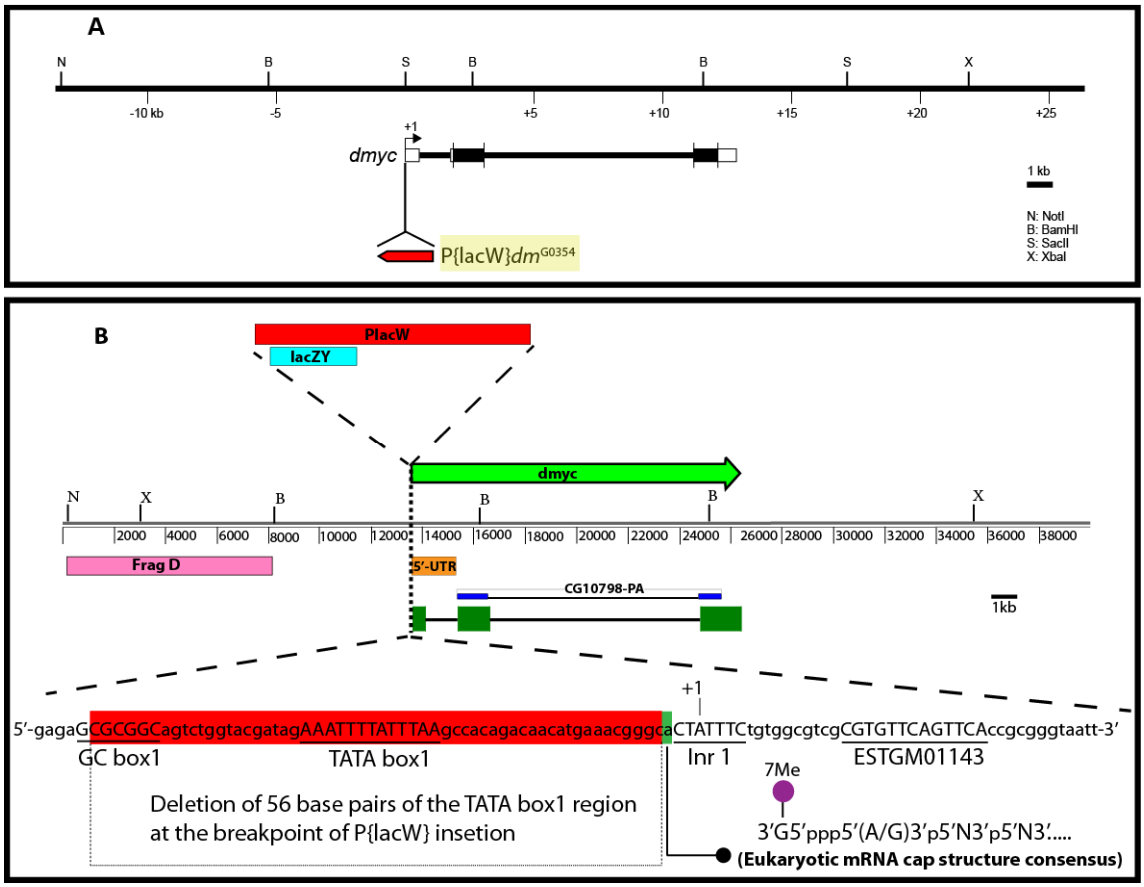



**Figure 2.6.d.** Insertion of the P-element *P{lacW}* at the TATA box1 region of the *dmyc* gene is lethal. (**A**) Insertion site of the reporter *lacZ* P-element (10.3-kb in size) at the *dmyc* locus is shown. Breakpoints of the insertion are as follows: 3D2, X:3267141..3267197, which maps to the region 213 nucleotides upstream of *dmyc* exon1 start site. (**B**) Insertion of the P-element *P{lacW}* causes deletion of 56 base pairs (red highlight), including GC box1 and TATA box1, at the breakpoint of insertion. A residue (green highlight adjacent to the breakpoint) was shown to be the first transcribed base (Figure 2) that favors capping (Latchman, David S. "Gene Control." 1st edition, New York: Garland Science, Taylor and Francis Group, LLC, 2010, page 176) of the *dmyc* mRNA.

In addition, it has been shown that the P-element *P{lacW}* is inserted into the endogenous *dmyc* promoter (Bourbon et al., 2002), and because the P-element insertion point is downstream of the TATA box, it disrupts transcription (Figure 2.6.c and Figure 2.6.d). This is not unusual for random P-element mutagenesis.

Investigation of cap structure in eukaryotic mRNAs has shown that in most RNAs the 7-methyl Guanine is linked through a 5' to 5' bond and three phosphate molecules to an A or G residue to be transcribed as first nucleotide (Latchman, David S. "Gene Control." 1st edition, New York: Garland Science, Taylor and Francis Group, LLC,

2010, page 176) (Figure 2.6.d). These results show that *dmyc* transcription may be initiated 24 bases downstream of TATA box1, an A residue at this locus being the first base to be transcribed. It further emphasizes that the immediate upstream elements in this region act as basal promoter elements for the assembly of general transcription factors.

## 2.7 Analysis of the *dmyc* Full-Length cDNA and Identification of the Splicing Signals in the *dmyc* Transcript

In eukaryotes, the noncoding intronic sequences of pre-mRNA molecules are removed through different splice mechanisms and the exons are joined together (Dogan et al., 2007; Qi et al., 2012). The recognition signals necessary for splicing involve the following *cis*-regulatory elements: branchpoint sequence, the polypyrimidine tract, and the 5' and 3' splice sites (Gao et al., 2008; Mount et al., 1992). Interaction of specific factors and RNAs with these cis-elements stepwise catalyze the formation of the spliceosome that via transesterification reactions remove the intervening sequences and ligate the exons (Wachtel and Manley, 2009; Will and Luhrmann, 2006). I performed 5' RACE analysis with gene specific primers, GSP-C05 and GSP-C06, that hybridize to the C-terminus exon and extends the *dmyc* mRNA from its 3'-end towards 5'-site (Figure 2.7.a).

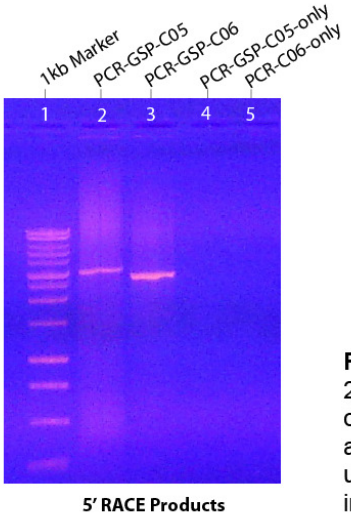



**Figure 2.7.a. Extension of the *dmyc* cDNA at its 3' end.** Lanes 2 and 3 show the polymerase chain reaction amplification products of the *dmyc* cDNA end with the primers GSP-C05 and GSP-C06. For all of the amplification reactions the Universal Primer Mix (UPM) was used as forward primer. Sequences for the oligonucleotides are listed in Table 5.1.

Analyses of the sequences at the ends of the introns confirmed that the first two nucleotides are GT whereas the last two bases are AG (Figure 2.6.c). We

hypothesized that like many eukaryotic mRNAs, the GU/AG nucleotide pairs should not appear in mature mRNA. Indeed, after RACE analysis of the *dmyc* cDNAs with different lengths we did not observe the GU/AG nucleotides in the mRNA, rather the ligation of the 3'-end of each upstream exon to the 5' end of the following exon in the splice junctions (Figure 2.7.b). From this result we conclude that *dmyc* largest transcript follows the constitutive and regulated splicing rules of GU/AG splicing mechanism to form the mature RNA for translation (Figure 2.7.c). Further we show that all three exons are detected in the transcript (Figure 2.7.b). This observation suggests that the exon1, despite its noncoding characteristic, might contain important translational signals for the synthesis of the *dmyc* protein. In adult *Drosophila* RNA mixture we could not detect a smaller transcript that could emerge through alternative splicing or transcription initiation from downstream promoter.

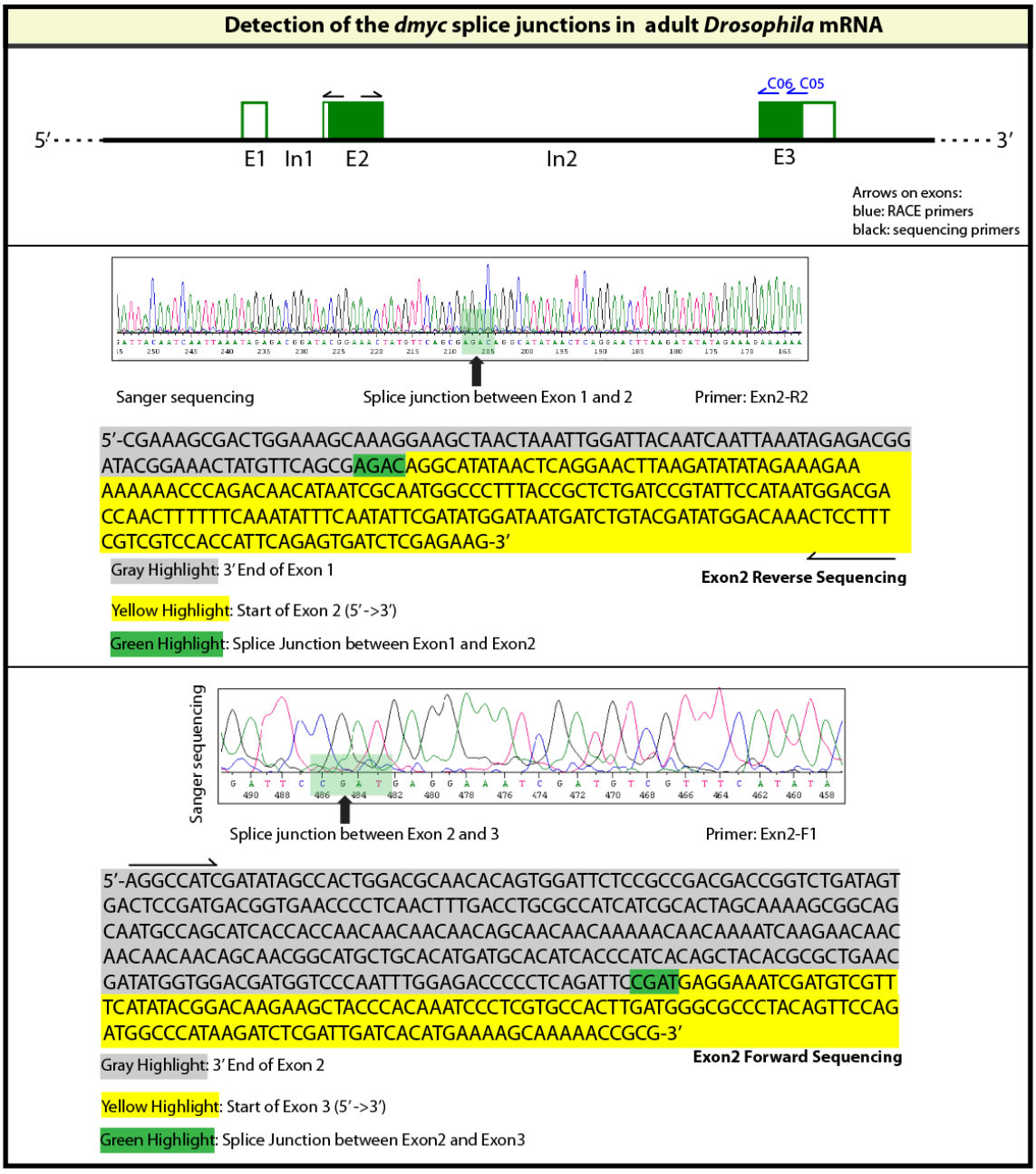
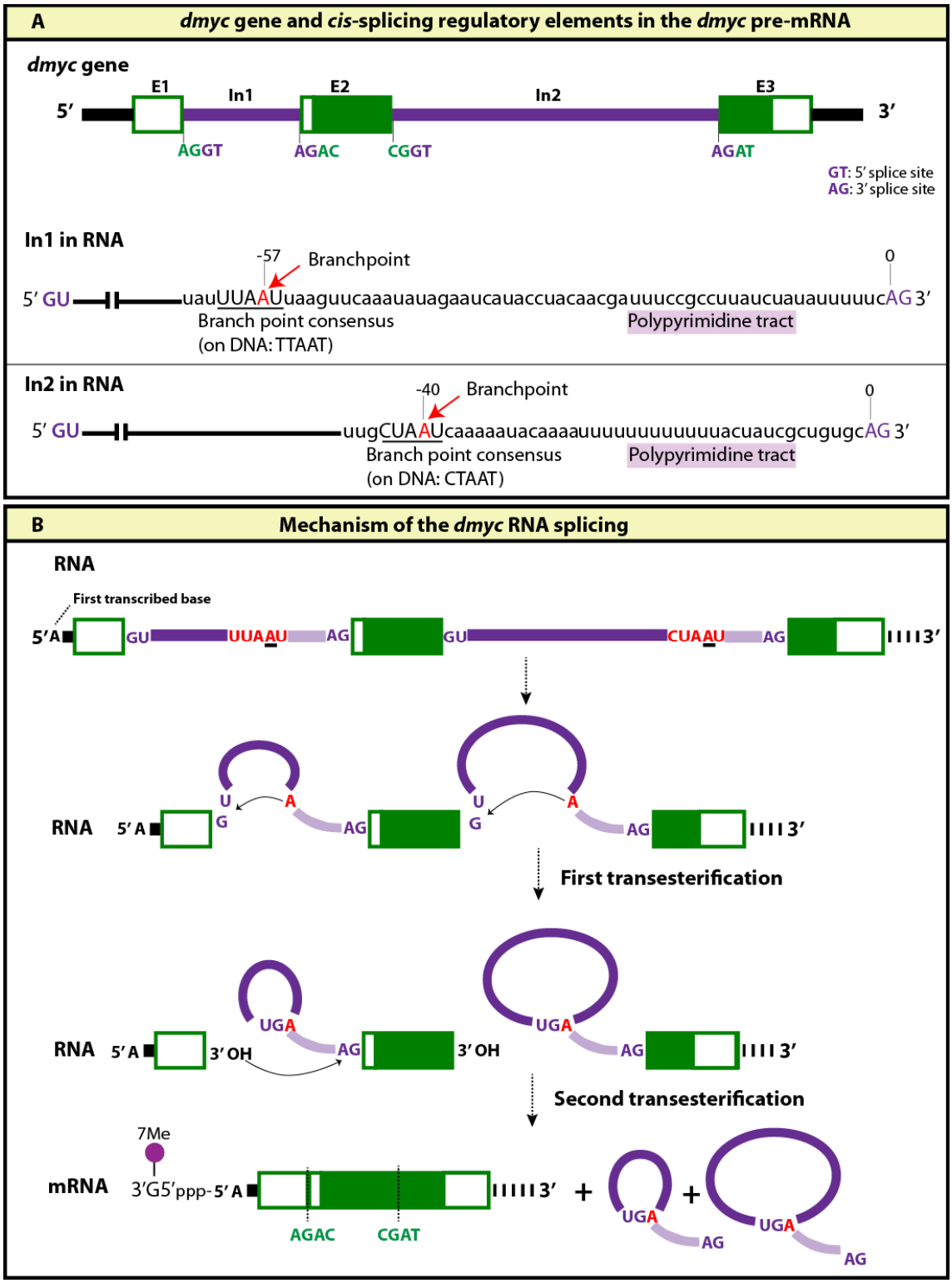



**Figure 2.7.b. 5' RACE analysis of *dmyc* cDNA ends determines splice junctions of the largest transcript.** *dmyc* genomic organization (not drawn to scale) is shown on top. Sanger sequencing results after analysis of the 5' RACE products that were amplified with the gene specific primers mapping to the exon3 (GSP-C05 and GSP-C06) are shown below. GSP-C05 and GSP-C06 hybridize to the *dmyc* exon3 and extend the full-length cDNA from its 3'-towards the 5'-end. The PCR products were sequenced with the reverse primer Exn2-R2 and the forward primer Exn2-F1 (black arrows). The PCR products were sequenced at Microsynth AG, Balgach, Switzerland. Splice junctions between exon1/exon2 and exon2/exon3 are indicated in green.


A
dmyc gene and cis-splicing regulatory elements in the dmyc pre-mRNA
dmyc gene
E1
In1
E2
In2
E3
5′
3′
AGGT
AGAC
CGGT
AGAT
GT: 5′ splice site
AG: 3′ splice site
In1 in RNA
-57
Branchpoint
0
5′ GU
uauUUAAUuaaguucaaauauagaaucauaccuacaacgauuuccgccuuaucuauauuuuucAG 3′
Branch point consensus
(on DNA: TTAAT)
Polypyrimidine tract
In2 in RNA
-40
Branchpoint
0
5′ GU
uugCUAAUcaaaaauacaaaauuuuuuuuuuuuuacuaucgcugugcAG 3′
Branch point consensus
(on DNA: CTAAT)
Polypyrimidine tract
B
Mechanism of the dmyc RNA splicing
RNA
First transcribed base
5′A
GU
UUAAU
AG
GU
CUAAU
AG
3′
RNA
5′ A
U
G
A
AG
First transesterification
RNA
5′ A
3′ OH
UGA
AG
3′ OH
Second transesterification
7Me
mRNA
3′G5′ppp-5′ A
AGAC
CGAT
UGA
AG
UGA
AG

Results

**Figure 2.7.c. Splice signals and splicing of the *dmyc* mRNA.** (**A**) *dmyc* genomic organization (not drawn to scale) with the nucleotides at the 5' splice donor and 3' splice acceptor sites (GT/AG in purple) are shown on top. Regulated splicing *cis*-elements of the *dmyc* RNA are shown below and include the following: 5' and 3' splice signals (GU/AG), branchpoint consensus sequence, and a polypyrimidine tract located adjacent to the AG sequence at the 3'-end of each intron. (**B**) The A residue (branchpoint) within the branchpoint consensus sequence participates in the formation of the ester bond between the 5' phosphate of the intron and the 2' oxygen of the branchpoint A residue to initiate the excision of the intron by building a lariat. Full excision of each intron requires two transesterification steps. Green nucleotides indicate exon/exon junctions.

# 3. Discussion

## 3.1 Computational Identification of Regulatory Units within the *dmyc* Locus

We have used multiple computational approaches to support the set-up of experimental procedures and interpret the findings obtained with the reporter activity studies. These approaches have helped to outline different aspects of the regulated patterning of *dmyc* throughout the development. Bioinformatics Tools used to track evolutionary conserved regions and other putative regulatory elements within noncoding sequences of the *dmyc* gene, and the conclusions drawn will be discussed.

### 3.1.1 Multiple Conserved Sequence Blocks

The majority of developmental genes achieve patterning via large noncoding regulatory regions containing numerous *cis*-regulatory elements and other diverse regulatory sequences (Carroll, 2008; Zeitlinger and Stark, 2010). Bioinformatics analyses of *c-myc* locus have yielded important insights into the molecular details of its regulation (Chen et al., 2007; Yap et al., 2011). Numerous *in vivo* and *in vitro* studies have shown that the computationally predicted elements regulate *c-myc* promoter in response to a variety of signals to integrate the dynamic inputs into the *c-myc* mRNA output (Chung and Levens, 2005; Liu and Levens, 2006).

Thus, we set out to determine whether the pattern of *dmyc* expression might be similarly regulated. Computational comparative searches of the 40-kb region spanning the *dmyc* gene by means of multialignment tool *Evorprinter* and subsequent analysis with *cis*-Decoder, detected multiple conserved sequence blocks common to most *Drosophila* species.

In the far upstream region we identified multiple conserved sequence blocks that could refer to possible regulatory motifs in this region. Indeed when testing fragment D for its ability to regulate reporter activity, it turned out to be active in presumptive neuromuscular tissues in late embryos (Figure 2.5.2.b). In proximal 5' region we identified potential regulatory motifs, including a conserved E-box. By removing the

conserved sequence blocks and the E-box in this region, the activity of *lacZ* reporter was abrogated in larval brain and imaginal tissues (transgene J7 in Figure 2.5.3.c)

In the 0.93-kb fragment of the large intron (J8.4 transgene in Figure 2.4.3.a and 2.5.4.a) we detected a wide range of multiple conserved sequence motifs being indicative of strong enhancer elements in this region. In addition to two conserved E-box sequences (CACGTG and CACTTG, Myc recognition motifs), the conserved cluster of (ATGTTGCCA), whose core (TGTTGC) is repeated three times, there is a dead-ringer for the HLHm-3-2 enhancer (CGCGTGGGAAAA, component of Notch-signaling), within which the consensus binding site (GTGGGAA) for *suppressor of Hairy su(Hw)* resides (Figure 2.1.1.b). While the full-length J8 transgene is active in *dmyc* manner in all tissues (Figure 2.4.3.a and 2.5.4.b), the enhancer region in J8.4 or the downstream promoter in J8.5, each separately, are inactive (Figure 2.4.3.a and 2.5.4.d). Like the *c-myc* promoter (Cleveland et al., 1988), *dmyc* has previously been shown to undergo autorepression. The identified E-box sequences may represent the basis for the previously demonstrated ability of dMyc to undergo autorepression to downregulate its transcription (Goodliffe et al., 2005).

To further confirm the evidence obtained with *cis*-Decoder for the existence of multiple binding sites in intron 2, we analyzed this region with the bioinformatics tool Tiffin (version 11.0) (Down et al., 2007) that uses the expression database related to transcription factor binding sites in the DNA sequence motifs to find putative transcription factor binding motifs (TFBMs). Consistent with our previous finding, the program predicts multiple binding motifs in the intron 2 region (Figure 3.1.1).

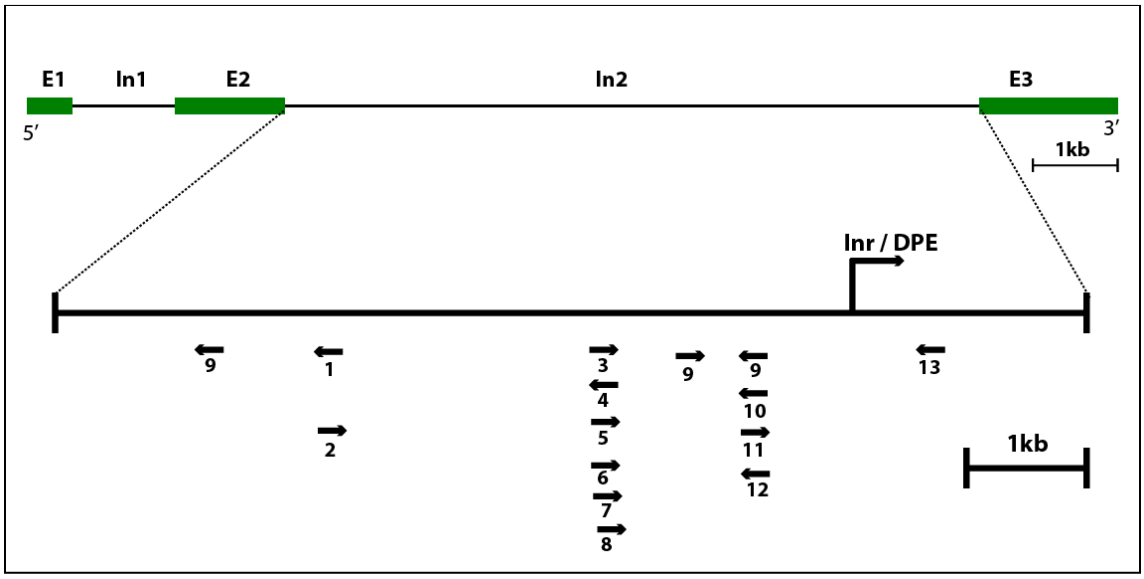



**Figure 3.1.1. Predicted Tiffin binding motifs (TFBMs) in the *dmyc* intron 2.** The motifs were predicted with the bioinformatics tool Tiffin Version 11.0 in the database for DNA sequence motifs. *dmyc* genomic organization and the intron 2 region are shown. The identified TFBMs are indicated with numbered arrows under the intron 2. The numbers indicate their relative distance from the 5'-end of the intron. The orientation of the sequence motifs is 5' → 3'. (E1: exon1; In1: intron1; E2: exon2; In2: intron2; E3: exon3).

Comparative analysis of the 3'-end of *dmyc* across 12 sequenced *Drosophila* species revealed multiple conserved sequence blocks in this region (Figure 2.1.1.b). Given that *c-myc* is regulated at the level of mRNA stability (Brewer, 1991) via conserved sequences in its 3'-UTR (Lemm and Ross, 2002), it will be of interest in the future to determine whether the stability of *dmyc* transcripts depends upon the presence of regulatory domains in its 3'-UTR. The conserved sequence blocks may contain potential microRNA target sites to serve for post-transcriptional modifications, as is the case for the majority of developmental control genes (Enright et al., 2003; Herranz et al., 2010). Indeed, *dMyc* has an active role in microRNA biology (Harris et al., 2011; Herranz et al., 2010), although regulation of *c-myc* by microRNAs has been reported (Aguda et al., 2008; Bueno and Malumbres, 2011; Bueno et al., 2008; Cannell and Bushell, 2010), the evidence for direct regulation of *dmyc* requires investigation. This work provides a starting point for investigating the putative microRNA binding sites and the mechanisms for the interactions between these motifs and their targets.

### 3.1.2 Prediction of Promoter Region within the 8-kb Far Upstream Sequences (Fragment D)

Combination of neural networks elements with genetic algorithms is one of the new approaches in recognizing potential promoter regions within eukaryotic genomic sequences (Knudsen et al., 2009). Initially, we had identified possible regulatory motifs and enhancer regions by mean of *Evo-Printer* / *cis*-Decoder in the *dmyc* far upstream fragment (Figure 2.1.1.a). Analysis of the fragment D in *dmyc-lacZ* reporter JD, detected an expression pattern restricted to late embryogenesis in body segments and in presumptive neuromuscular tissues (Figure 2.5.2.b). A search with neural network genetic algorithm Promoter 2.0, revealed a potential promoter region in these sequences that may be recognized by Pol II. Regulation by these *cis* elements may be required during embryogenesis, where dMyc is required to specify neuronal fate and facilitate neuroblast proliferation (Orian et al., 2007) and in control of mesodermal fate determination (Demontis and Perrimon, 2009). In light of this finding, further analysis of *dmyc* transcriptional regulation in this region in response to developmental signals will be of great interest.

### 3.1.3 Prediction of TATA Box

Analysis of the 5'-d*myc-lacZ* deletion construct, containing intron 1, the 5'-UTR, and 100 bp upstream of the predicted transcription start site (J7), revealed that this minimal region was sufficient to give reporter activity in a *dmyc*-like pattern in both ovarian nurse cells and in the embryo, but not in larval tissues (Figure 2.5.3.c). Therefore, we inspected the region extending from nucleotide 100 upstream of the 5'-UTR to nucleotide +187 for initiator consensus sequences. In most mammalian protein-coding genes, there is a TATA box located 25–30 bp upstream of the transcription start site, an initiator element (Inr) overlapping the start sites (Breathnach and Chambon, 1981; Juven-Gershon et al., 2006; Kollmar and Farnham, 1993; Smale and Kadonaga, 2003), and/or a GC-box (SP1 binding site) 60–100 nucleotides upstream of the transcription start site (Black et al., 2001; Lang et al., 1988; Liu et al., 2008). Experiments with vertebrate cell lines and *Drosophila* embryonic extracts have revealed strict conservation of the Inr consensus sequence, Py Py $A_{+1}$ N T/A Py Py among vertebrates and invertebrates (Lo and Smale, 1996; Smale and Kadonaga, 2003).

Analysis of J7 (the region 1914 bp upstream of the predicted translation start site) and the expressed sequence tag (ESTGM01143; beginning at 1812 bp upstream of the translation start) with Lasergene GeneQuest module revealed that the 102 bp sequence between the 5'-end of the expressed sequence tag and the 5'-end of the J7 genomic sequence contains a perfect Inr consensus sequence, a TATA box, and a GC box (GC box1/TATA box1/Inr1 Figure 3.1.3). The TATA box is located 39 bp upstream of $A_{+1}$ in the initiator element. Previous reports have shown that occurrence of a TATA box 25–30 bp upstream of the Inr in the same core promoter leads to cooperation between the two elements to enhance promoter strength (Smale and Kadonaga, 2003). Although the distance of 39 bp in the *dmyc* promoter is on the edge of this optimum, cooperation between the TATA box and the Inr element has been shown to extend up to 90 bp in yeast promoters (W Schaffner, personal communication).

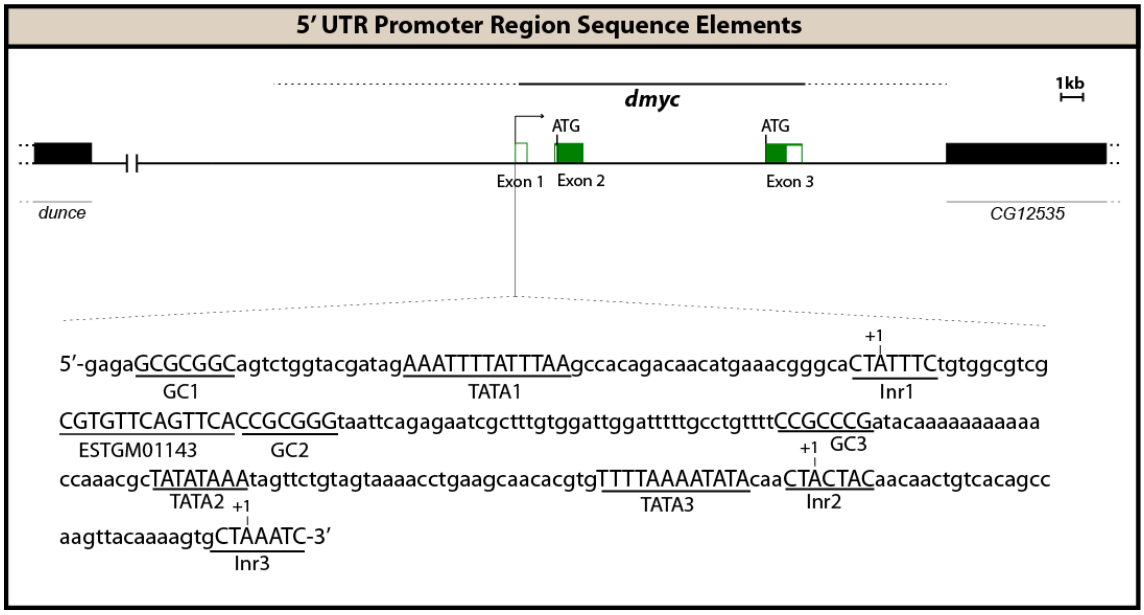



**Figure 3.1.3. Prediction of transcription start site in the *dmyc* 5'-UTR.** Structure of the *dmyc* gene and the two major translation start sits (ATG, ATG) are indicated on the top. The sequences of the major elements for the promoter in the 5'-UTR region, shown at the bottom, were identified by search with Lasergene GeneQuest module (GC1, 2, 3: SP1 binding site; TATA1, 2, 3: TATA box region; Inr: initiator element; ESTGM01143: known expressed sequence tag).

In addition to the TATA box1, Inr1, and GC box1, two putative TATA boxes (TATA2, TATA3), two GC boxes (GC2, GC3), and two Inr elements (Inr2, Inr3), as shown in Figure 3.1.3, were identified 180 bp downstream of the expressed sequence tag start site. The putative Inr2 element shows one deviation from the Inr consensus sequence at position +4, but the critical positions for Inr activity are +1 and +3, and single bp substitutions at the -2, -1, +4, and +5 positions can still produce an active Inr (Javahery et al., 1994), suggesting the second Inr element might be functional. The putative Inr3 element shows no deviation from the Inr consensus sequence, suggesting that Inr3 might be as functional as Inr1. TATA box2 is located 57 bp upstream of $A_{+1}$ in the Inr2 element and TATA box3 is located 55 bp upstream of $A_{+1}$ in the Inr3 element, both distances less than 90 bp in yeast promoters.

Later, by experimental investigation of transcription initiation site from this region (see Results, section 2.6), it turned out the TATA box1 is utilized to initiate *dmyc* transcription for its largest transcript (see Discussion, section 3.3.1).

### 3.1.4 Prediction of Downstream Promoter Element (DPE)

In support of the findings for *lacZ* reporter activity under control of intragenic region (Figure 2.5.4.b), we searched for Inr and downstream promoter elements, a sequence motif common to all *Drosophila* downstream promoters, in this region. Most protein-coding genes of *Drosophila* contain a downstream core promoter element that functions cooperatively with an initiator to facilitate the binding of transcription factors in the absence of a TATA box (Burke et al., 1998; Kutach and Kadonaga, 2000). A search for consensus *Drosophila* Inr (T-C-$A_{+1}$-G/T-T-T/C) and downstream core promoter elements (A/G-G-A/T-C-G-TC) (Figure 3.1.4) using DNASTAR Lasergene 9.1, GeneQuest Module, revealed the presence of downstream promoter sequence motifs comparable with the *Drosophila* consensus Inr/downstream core promoter element (Kutach and Kadonaga, 2000). Sequence motifs were typed in GeneQuest "type in pattern function" and searched for with a threshold of 100% (no errors allowed). The Inr and downstream promoter element motifs in the distal 2.1-kb of the intron 2 DNA sequence (Figure 3.1.4) met all the strict criteria for such elements, in that the Inr sequence motif (T-C-$A_{+1}$-T-T-C) does not deviate from the consensus, the downstream core promoter element (G-G-T-C-G) is identical to the

core consensus, the spacing between the downstream core promoter element and the Inr (34 nucleotides) is appropriate, and a G nucleotide is correctly positioned between the Inr and the downstream core promoter element (Figure 3.1.4).

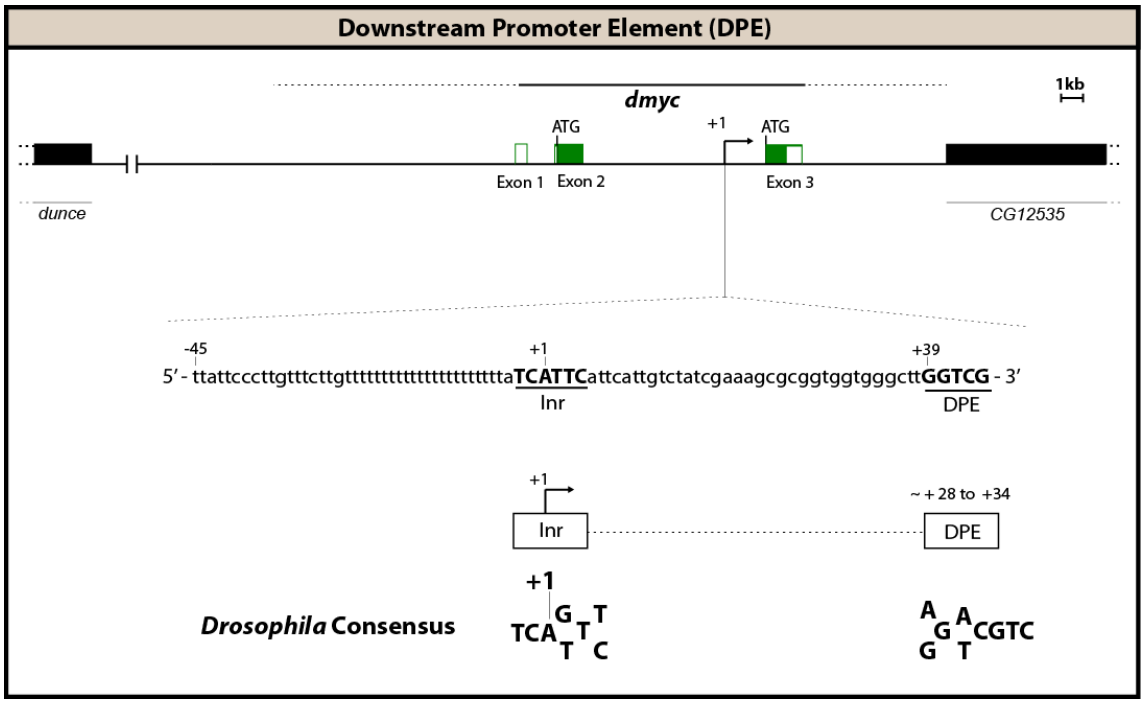



**Figure 3.1.4. Depiction of the *dmyc* genomic locus and the predicted downstream promoter elements Inr and DPE in the large intron.** The identified sequences for Inr and DPE in the *dmyc* intron2 are underlined, and the *Drosophila* consensus sequences (Burke et al., 1998) for the Inr and DPE are indicated at the bottom.

### 3.1.5 Prediction of Polyadenylation Signals in the *dmyc* 3'-UTR

The sequences between nucleotide -554 upstream of the translation stop to nucleotide +9763 downstream of the translation stop were searched with Lasergene 9.1 GeneQuest module for possible poly-adenylation signals and poly-adenylation sites. In animal cells, there is a hexanucleotide-sequence of AAUAAA (occasionally AUUAAA) that is located 10–35 nucleotides upstream of the polyadenylation signal (Anderson et al., 1981; Zhao et al., 1999). In addition, there is a U/GU-rich region (cleavage stimulation factor binding site) 14–70 nucleotides downstream of the polyadenylation signal (Lutz, 2008). We searched for polyadenylation signals in the *dmyc* 3'-end comparable with the consensus sequence using the GeneQuest module with a threshold of 100% (no errors allowed). The threshold for the T-rich region was

set at 80%. Analysis with the AATAAA sequence identified two potential polyadenylation signals (poly (A)1 and poly (A)3) (Figure 2.5.6.a and Figure 3.1.5), highly similar to the polyadenylation consensus sequence found in animal cells (Anderson et al., 1981). A search for ATTAAA sequences on the DNA sense strand detected poly (A)2, with a high degree of homology to the consensus sequence (Figure 2.5.6.a and Figure 3.1.5). In order from proximal to distal, we named them poly (A)1, poly (A)2, and poly (A)3 (Figure 3.1.5).

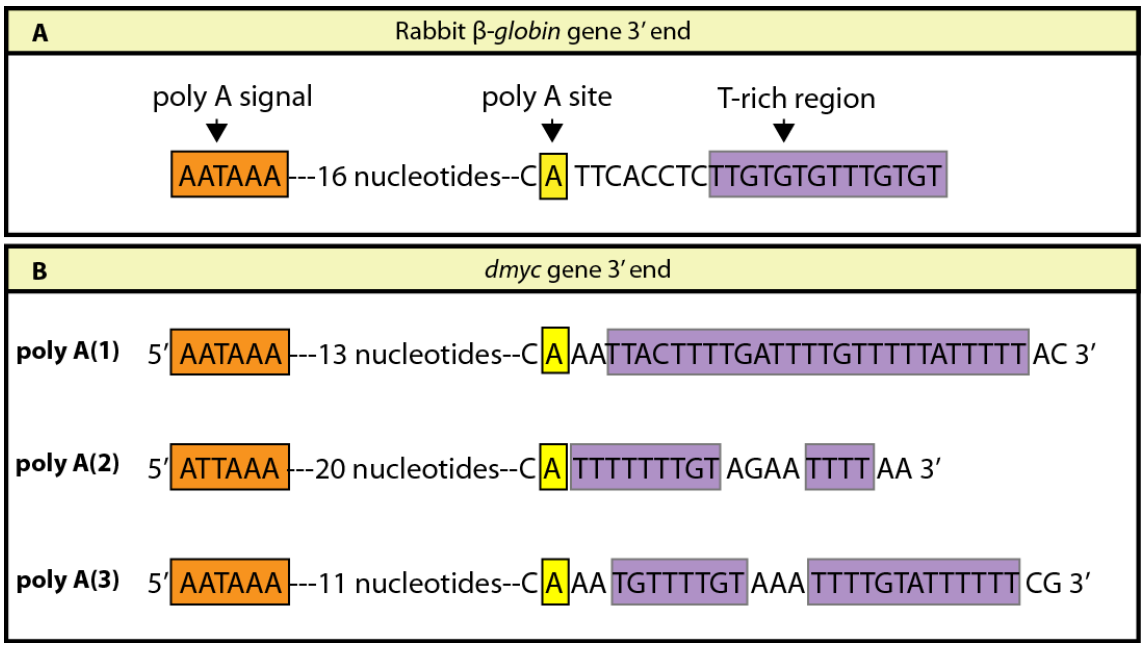



**Figure 3.1.5. Mammalian polyadenylation site consensus sequence and potential polyadenylation sites in the *dmyc* gene.** (**A**) Rabbit β-*globin* gene 3'-end is depicted as an example for a consensus polyadenylation signal. (**B**) 3'-end of the *dmyc* gene with the predicted polyadenylation signals is shown. The first A at the 5'-end of the polyadenylation signal, poly (A)1, corresponds to nucleotide +3761 downstream of the 3'-UTR, the first A at the 5'-end of poly (A)2 corresponds to nucleotide +5245 relative to the 3'-UTR, and the first A at the 5'-end of poly (A)3 corresponds to nucleotide +5952 downstream of the 3'-UTR.

## 3.2 *lacZ* Patterning under Control of *dmyc* Regulatory Regions

Computational comparative searches of the 40-kb region spanning the *dmyc* gene, detected multiple conserved sequence blocks. Our subsequent analysis of the *dmyc-lacZ* reporter constructs, containing all of the conserved sequence blocks (Figure 3.2), suggested that they were transcriptionally active and generated similar patterns of reporter activity as that described for the endogenous *dmyc-lacZ* enhancer trap. In

the following sections I shall discuss *lacZ* patterning under control of the *dmyc* noncoding regions.

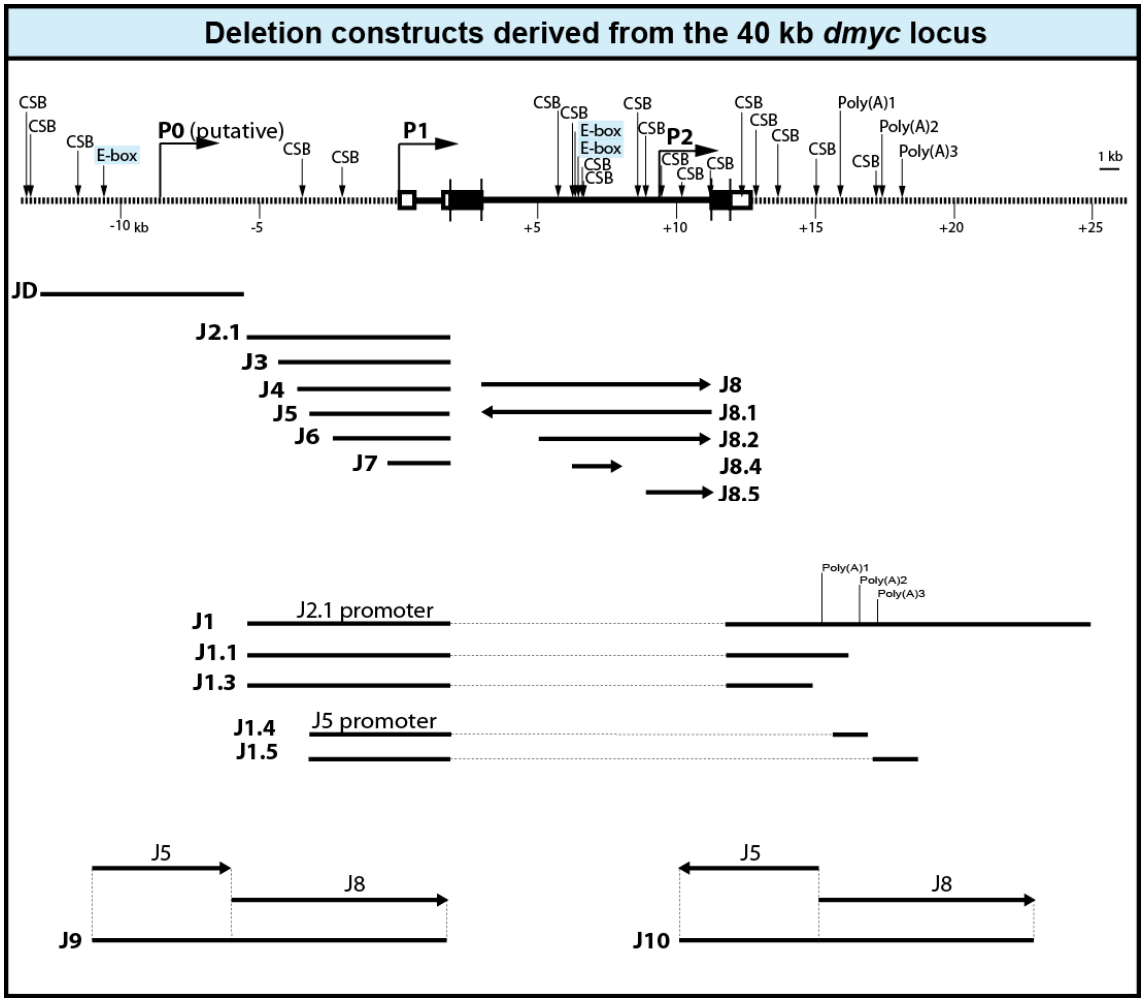



**Figure 3.2. Summary of the computational analyses of the *dmyc* locus and reporter constructs derived from the noncoding regions.** Top, the 40-kb *dmyc* locus was scanned with *EvoPrinter* and *cis*-Decoder to detect multiple conserved sequence blocks and E-boxes within the noncoding DNA sequences. The predicted promoter regions P0 (putative), P1, and P2 are shown. The scanning with Lasergene GeneQuest module detected three potential polyadenylation signals [Poly (A)1, Poly (A)2, Poly (A)3]. Bottom, deletion constructs derived from the 40-kb locus and their names are given.

### 3.2.1 Reporter Activity Regulated by Far Upstream Sequences

*lacZ* reporter for the far upstream 8-kb fragment produced an expression pattern restricted to late embryogenesis in body segments and in presumptive

neuromuscular tissues (Figure 2.5.2.b). *In silico* analysis revealed possible regulatory motifs in this region, including core promoter elements and conserved sequence blocks. Regulation by these *cis* elements may be required during embryogenesis, where *dMyc* is required to specify neuronal fate and facilitate neuroblast proliferation (Orian et al., 2007) and in control of mesodermal fate determination (Demontis and Perrimon, 2009). In light of this finding, further analysis of *dmyc* transcriptional regulation in this region in response to developmental signals will be of great interest.

### 3.2.2 *lacZ* Patterning under Control of 5'-UTR Sequences

Our analysis of the *dmyc-lacZ* reporter constructs, containing all of the conserved sequence blocks, suggested that they were transcriptionally active and generated similar patterns of reporter activity as that described for the endogenous *dmyc-lacZ* enhancer trap line (Figure 2.5.3.b and Figure 2.5.3.c/J2.1-J6). Analysis of the 5'-d*myc-lacZ* deletion construct, containing intron 1, the 5'-UTR, and 100 bp upstream of the predicted transcription start site (Figure 2.5.3.c/J7), revealed that this minimal region was sufficient to give reporter activity in a *dmyc*-like pattern in both ovarian nurse cells and in the embryo, but not in larval tissues. Characterization of binding sites in the tissue specific enhancers in this region and factors that bind to these elements will be of great interest.

### 3.2.3 *lacZ* Patterning under Control of Intragenic Region

Analysis of the *dmyc* large intron (Figure 2.5.4.b) shows a dynamic patterning of *lacZ* activity during early stages of development (Kharazmi et al., 2011), as is the case for most of developmentally regulated introns in *Drosophila* transcriptome (Gasch et al., 1989; Roy et al., 2010). Indeed, previous work has suggested that both *Drosophila* (Neto-Silva et al., 2010) and mammalian (Wierstra and Alves, 2008) *myc* transcription is also regulated via intronic promoter sequences. In support of these findings, we demonstrated that the full-length J8 transgene and its truncation J8.2, which contain just the intron 2 sequence of the *dmyc* gene, result in *lacZ* reporter activity in all tissues examined (Figure 2.5.4.b). Thus we searched for Inr and downstream promoter elements, a sequence motif common to all *Drosophila* downstream promoters in this region, and identified a DPE element in this region

(Figure 3.1.4). Additionally, Intron 2 contains multiple conserved sequence blocks and two conserved E-boxes (Figure 3.2). Deletion of the multiple conserved sequence blocks, upstream of the DPE element, abrogates reporter activity in virtually all the examined tissues (Figure 2.5.4.d). In light of this finding, analysis of RNA extracts taken from *Drosophila* ovary or embryonic tissues for the identification of a shorter form of *dmyc* protein will be of great interest. The knowledge gained here, combined with further molecular genetic approaches, will be informative for understanding the regulation of the *dmyc* from this initiation site.

### 3.2.4 Transcription Termination under Control of the *dmyc* 3'-UTR

Post-transcriptional 3'-end formation or polyadenylation of the mRNA precursor is a crucial step in mRNA maturation, in which most eukaryotic mRNAs acquire a poly (A) tail at their 3'-ends to promote transcription termination (Zhao et al., 1999), transport of the mature mRNA from the nucleus (Huang and Carmichael, 1996), and to enhance the translation and stability of mRNA (Preiss and Hentze, 1998). Analysis of the entire *dmyc* 3'-UTR for polyadenylation signals and polyadenylation sites revealed three potential consensus sequences, i.e., poly (A)1, poly (A)2, and poly (A)3 (Figure 2.5.6.a and Figure 3.1.5), which are all capable of terminating transgene transcription (Figure 2.5.6.b and Figure 2.5.6.c). In addition, the 4.362-kb DNA sequence upstream of poly (A)1 at the *dmyc* 3'-end leads to reporter activity in the pattern predicted for *dmyc* expression. Therefore, the *dmyc* gene would be predicted to produce different transcripts with shorter and longer lengths, consistent with the previous analysis of *dmyc* mRNA in which genomic probes derived from this region revealed three alternative transcripts with the lengths of 4.5-kb, 5-kb, and 6-kb (Gallant et al., 1996).

## 3.3 Transcription Initiation from P1 Promoter

In eukaryotic cells transcription initiation is highly complex and requires the assembly of general transcription machinery at core promoter prior to RNA polymerase II (Pol II) binging (Gross and Oelgeschlager, 2006). Analysis of core promoter elements and identification of initiation sites is a prerequisite to understanding the mechanism of transcription control (Sandelin et al., 2007).

In the study reported here, we have investigated the initiation site of the P1 promoter in the *dmyc* 5'-UTR. First, using a gene specific primer that hybridizes to *dmyc* exon 1, we performed 5' RACE experiments to synthesize the 5' portion of the *dmyc* cDNA that subsequently served as template to amplify the fragment. Our results indicate that only one initiation site is detectable in this region, and the core promoter consists of GC box1, TATA box1, and Inr1. This observation confirms our former computational prediction of the above regulatory elements in the 5'-UTR (Kharazmi et al., 2011). Examining of a *dmyc* haplo-lethal fly stock ($w^{67c23}$ *P{lacW}*$dm^{G0354}$/FM7c; Bloomington Stock Number 11981) (Bourbon et al., 2002), showed that the inserted P-element has broken away the TATA region; as a result the affected copy has become nonfunctional. Further, the finding is consistent with the fact that most eukaryotic promoters consist of a TFIIB Recognition Element (BRE, immediately upstream of TATA box), a TATA box, and an initiator element that is located 25 to 30 bp downstream of the TATA box (Smale and Kadonaga, 2003). The number and the relative distance of the core promoter elements appear to be appropriate to confer *dmyc*, as immediate early gene, strength for the action of its promoter. This work provides a starting point for investigating the exact sequences of each element and the assembly of general transcription machinery.

### 3.3.1 Mechanism of Splicing of the *dmyc* Largest Transcript

In higher eukaryotes, the removal of introns from pre-mRNA molecules is mediated through a complex process that involves many factors. The splice *cis*-elements participate in two transesterification reactions performed by a complex structure known as spliceosome to remove the introns and ligate the upstream exon to the downstream exon (Figure 2.7.c). Indeed analysis of the *dmyc* full-length cDNA for splice signals in the transcript, revealed that like virtually all eukaryotic introns, both *dmyc* introns include a 5' and a 3'-end splice signal as well as a branchpoint consensus and a polypyrimidne tract located adjacent to the AG sequence within each intron (Figure 7A, B, C). The branchpoint consensus sequence in *Drosophila* resembles: 5' CTA<u>A</u>T 3', and occasionally 5' TTA<u>A</u>T 3', in which branch formation occurs at underlined A (Mount et al., 1992). The consensus sequence is present in intron 2 whereas intron 1 contains the less conserved sequence (Figure 2.7.c). Sequencing of the amplified cDNA ends showed that the *dmyc* introns are spliced out

following the GU/AG rules in constitutive and regulated splicing to accomplish the two transesterification reactions. Consequently, exon 1 is ligated to exon 2 and exon 2 to exon 3 so that the full-length mRNA contains all three exons. Indeed the dMyc protein is 717 amino acids, while c-Myc is about 459 amino acids and both proteins contain domains that are required for Myc function (P Bellosta, personal communication). In the study here, we have shown the splice mechanism for the full-length *dmyc* transcript. However, we can't rule out the possibility of alternative splicing during early stages of development that may produce shorter transcripts.

## 3.4 Conclusions and Perspectives

The dynamic expression of *dmyc* is initiated from multiple transcription start sites, as summarized in Figure 3.4.a and Figure 3.4.b. Tight regulation of dMyc is crucial for cell growth and division during the early phases of development and cell fate specification.

| | Reporter Activity in Tested Tissues | | | | | | |
|---|---|---|---|---|---|---|---|
| Tissue / Construct | Brain | Wing | Eye | Leg | Early Embryo | Late Embryo | Ovary |
| JD | No | No | No | No | No | Yes | No |
| J2.1 | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| J3 | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| J4 | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| J5 | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| J6 | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| J7 | No | No | No | No | Yes | Yes | Yes |
| J8 | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| J8.1 | No | No | No | No | No | No | No |
| J8.2 | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| J8.4 | No | No | No | No | No | No | No |
| J8.5 | No | No | No | No | No | No | No |
| J9 | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| J10 | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| J1- J1.5 | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

**Figure 3.4.a. Reporter activity under control of the *dmyc* noncoding regions.** The table shows the summary of the activity of *lacZ* reporter constructs in the tested tissues (for details see Results, section 2.5).

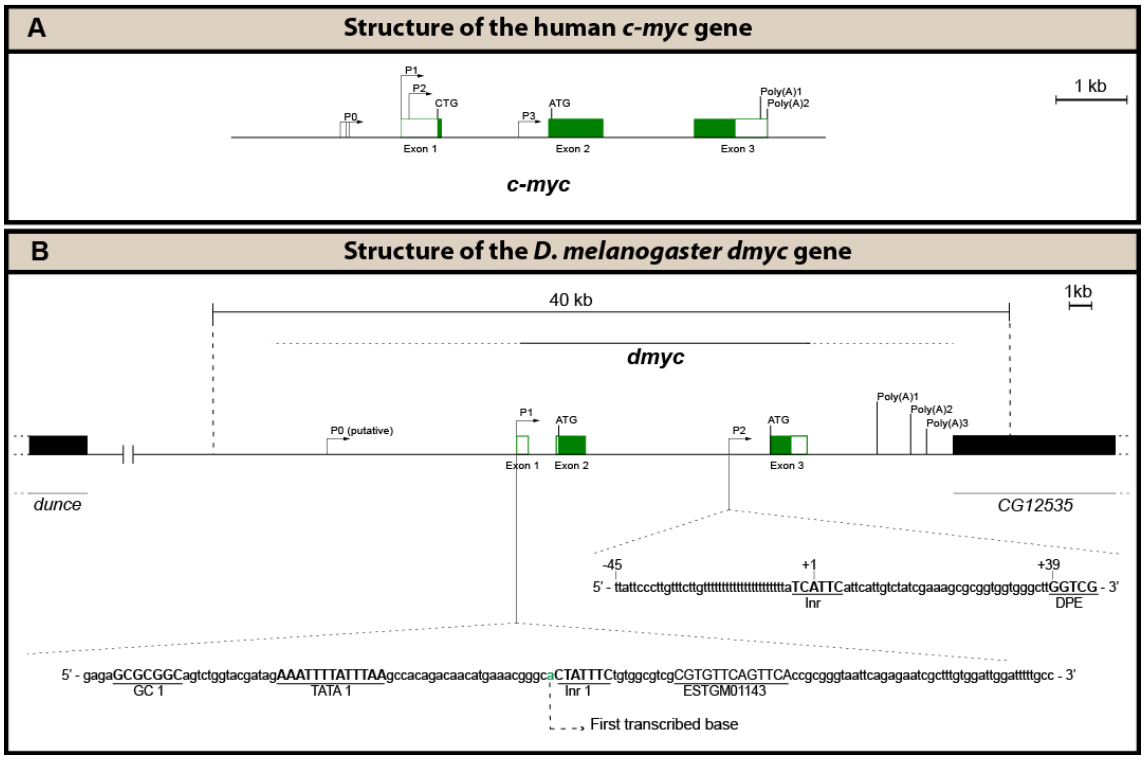



**Figure 3.4.b. During development, the activity of the *dmyc* regulatory region is integrated by multiple transcription start sites.** (**A**) Human *c-myc* gene contains three exons (white boxes indicate noncoding regions, the green boxes represent coding regions), four promoters (P0, P1, P2, P3), two major translation start sites (CTG, ATG) and two polyadenylation signals [Poly (A)1, Poly (A)2]. (**B**) *Drosophila myc* gene contains three exons (white boxes noncoding, the green boxes coding regions), two main promoters (P1, P2), one putative promoter (P0), two major translation start sites (ATG, ATG), and three potential polyadenylation signals (Poly (A)1, Poly (A)2, Poly (A)3). The specified regulatory elements are plotted on the basis of the experimental data and the computational analyses of the *dmyc locus*. The sequences of the major elements for the promoters P1 and P2, shown at the bottom, were identified by search with Lasergene GeneQuest module.

Because c-Myc is a potent mitogen, the level of *c-myc* transcription must be tightly regulated. Myc transcription responds to developmental signaling molecules (de la Cova et al., 2004; Duman-Scheel et al., 2004; Johnston and Edgar, 1998), which are likely to modulate the complement of a wide variety of transcription factors at the *myc* promoter (Hallikas et al., 2006; Liu et al., 2006). The evidence presented in this work reinforces the idea that *dmyc* represents a tightly and dynamically regulated gene. Further genetic studies combined with genomic approaches will be required to identify the molecular mechanism controlling *dmyc* transcription via the regulatory elements identified here.

# 4. Materials and Methods

## 4.1 Flies

Flies were maintained in ½ l synthetic bottles and synthetic fly husbandry vials on a standard yeast, cornmeal, sugar, nipagin, and agar medium at 18°C and 25°C in a light/dark 12h/12h photoperiod.

***dmyc-lacZ* P-Elements:** The established *dmyc-lacZ* transgenic flies carry a transposon containing the *E. coli lacZ* cDNA under the control of different fragments from the *dmyc* noncoding regulatory regions. The transposon contains a *white*$^{+}$ minigene as a selectable marker. The *dmyc* regulatory regions are preceding the *lacZ* cDNA (see Appendix, section 5.3).

## 4.2 Bioinformatics Analyses of *dmyc* Locus

For defining the genomic organization of the *dmyc* gene in twelve sequenced *Drosophila* species, the bioinformatics tool DNASTAR Lasergene 9.1 MegAlign module was used to edit the sequences taken from FlyBase. We used the Lasergene 9.1 MegAlign module with the settings “Multiple Alignment, ClustalW” (Thompson et al., 1994). The neural network genetic algorithm PROMOTER 2.0 was used to predict promoter regions in the far upstream sequences using CCAAT or bHLH recognition motifs (Knudsen, 1999). The *dmyc* 5'-end was searched with the DNASTAR Lasergene 9.1 GeneQuest module for prediction of TATA box-Inr elements, the intron 2 region for the existence of an Inr-downstream core promoter element, and the 3'-end was searched for the prediction of polyadenylation sequence motifs. The phylogenetic footprinting tools, *EvoPrinter* and *cis*-Decoder (Brody et al., 2007; Yavatkar et al., 2008) were used to detect E-boxes, bHLH binding sites, and multiple-conserved sequence blocks in the *dmyc* region common to most *Drosophila* species.

## 4.3 Cloning of the *dmyc* Locus from the BAC Clone RP98-2A13

The *Drosophila* BAC clone RP98-2A13, contains a ~176-kb genomic fragment from the *Drosophila* X chromosome. The BAC vector (9176bp), into which the genomic fragment has been cloned, is called pBACe 3.6 and contains chloramphenicol as

resistant gene. The *dmyc* locus lies at the beginning of the 5'-end of the genomic fragment in the BAC vector. The BAC clone was received as stab culture. After isolating a single correct colony and the BAC DNA preparation with Qiagen large construct kit, the DNA was digested with Not I-XbaI-DraIII simultaneously to isolate the insert RP27, a 27-kb Not I-XbaI fragment harbouring the *dmyc* locus of ~ 12.6-kb. The fly transforming vector pC-Asp718 was linearized with NotI-XbaI at its MCS, dephosphorylated with CIAP (Roche Scientifics), ligated with the insert RP27 to generate pC-RP27. The plasmid pC-RP27 served as a substrate to isolate the 5'-end sequences of the *dmyc* gene:

pC-RP27 = pC-Asp718 x NotI x XbaI + Insert: RP98-2A13 x NotI x XbaI (27-kb band).

The BAC clone RP98-2A13 was sequenced at the beginning and the end of the *dmyc* gene, each time toward the *dmyc* gene, with the primers BAC-F and BAC-R (Table 5.1; No. 3 and No. 4). The plasmid pC-RP27 was sequenced with primers pcaF and pcaR at the subcloning junctions (Table 5.1; No. 9 and No. 10).

## 4.4 Generation of *lacZ* Reporter Strains

In the following sections I shall explain the creation of reporter constructs for studying *dmyc cis*-elements. For random P-element transformation, the ready-to-use transforming vector pCaSpeR-NLSlacZ (Drosophila Genomic Resources, originally from Tummel's laboratory) was used. The pCaSpeR4-NLSlacZ vector, free from hsp70 promoter or any other regulatory elements, contains the eye marker *white* gene, the reporter NLSlacZ, SV40 poly (A) tail, and the ampicillin resistance gene.

For site-specific integration the pattB-temp del LoxP reporter plasmid (Rainbow Transgenic Flies, original pUAST-attB vector was a gift from Basler's laboratory) was used. The original vector pUAST-attB was engineered to remove the 5×UAS-hsp70 and LoxP site sequences, followed by self-ligation to obtain the ready-to-use pattB-temp del LoxP transforming vector. The pattB-temp del LoxP reporter plasmid also contains the eye marker *white* gene, the reporter NLSlacZ, SV40 poly (A) tail, and the ampicillin resistance gene.

### 4.4.1 Construction of Proximal and 5'-UTR Transgenes

Inserts for the reporter constructs, J2.1–J7 are derived from the genomic sequences in pC-RP27. PC-RP27 was digested with BamHI and run on 0.5% agarose gel to isolate the 7.9-kb BamHI-BamHI *dmyc* 5' fragment. The BamHI-BamHI fragment was subcloned into pBS-SKII (+) (which was a generous gift from Oleg Georgiev, Walter Schaffner's laboratory) to obtain SKII-dmyc5 intermediate plasmid. The *dmyc* 5' 7.9-kb BamHI-BamHI fragment in SKII-dmyc5, was first mutated by polymerase chain reaction to remove the approximately 800 bp open reading frame sequence and to introduce an Acc65I restriction site at the 3' end of the fragment. For polymerase chain reaction amplification of the *dmyc* promoter/enhancer, the BAC clone RP98-2A13 served as the template, the amplified fragment was 2601 bp in size, and the amplifying primers had the names dm5E2F and dm5E2R (Table 5.1). Polymerase chain reaction conditions in the thermal cycler were one cycle of initial denaturation at 98℃ for 30 seconds; 30 cycles of denaturation at 98℃ for 10 seconds, annealing at 60℃ for 30 seconds, and extension at 72℃ for 60 seconds; and one final extension cycle at 72℃ for 5 minutes, held at 4℃. Plasmids SKII-dmyc5 and the polymerase chain reaction fragment were each separately digested with AatII and Acc65I (AatII is located 1550 bp downstream of the transcription start and Acc65I is in the SKII-dmyc5, 60 bp downstream of the *dmyc* fragment in the multiple cloning site). The vector SKII-dmyc5 was digested to excise an approximately 1150 bp AatII/Acc65I fragment. The polymerase chain reaction fragment was digested to isolate the approximately 296 bp AatII/Acc65I fragment. After isolation on 1.5% agarose gel, the vector and insert were ligated over-night to create plasmid SKII-dmyc5-ORF, free from open reading frame (10,035 bp in size). The 5' NotI-Acc65I 3' *dmyc* 5' fragment in the plasmid SKII-dmyc5-ORF, was excised and subcloned in 5' NotI-Acc65I 3' linearized/dephosphorylated pCaspeR4-NLSlacZ (pC-NLSlacZ) in front of reporter *lacZ* to obtain the transgene J2.1.

To create the J4, J5, and the J7 deletions, J2.1 was digested with NotI and another enzyme (see below). That other enzyme could cut more than one time but it did not cut in the remaining fragment. The Klenow treatment was carried out by standard molecular biology techniques (Rubin, 1988). Self-ligation took place at 16°C in a big volume of reaction mix (200µl instead of 10-20µl conventional volume for a ligation reaction) to increase the chance for intermolecular interactions (self-ligation).

For creating J3 and J6, the deletions were first done in SKII-dmyc5-ORF using XbaI to cut at the 5' end and MscI or NarI respectively to cut internally to create the template vectors J3-temp (8370 bp) and J6-temp (5459 bp). J3-temp and J6-temp were sequenced at NotI junction with the M13 universal reverse primer. In a further step, the shorter NotI-Asp718 fragments in J3-temp and J6-temp were substituted for the NotI-Asp718 fragment in J2.1:

J2.1 = pC-NLSlacZ X NotI X Asp718 + Insert: SKII-dmyc5-ORF X NotI X Asp718 (7.1-kb band)

J3 = J2.1 X NotI X Asp718 + Insert: SKII-dmyc5-ORF X MscI X XbaI X Klenow X NotI X Asp718 (1.6-kb band)

J4 = J2.1 X NotI X NheI X Klenow

J5 = J2.1 X NotI X SpeI X Klenow

J6 = J2.1 X NotI X Asp718 + Insert SKII-dmyc5-ORF X NarI X XbaI X Klenow X NotI X Asp718 (4.6-kb band)

J7 = J2.1 X NotI X SphI X T4.

J2.1-J7 constructs were sequenced at the subcloning junctions using the primers pZR and pcaF (Table 5.1; No. 8 and No. 9).

### 4.4.2 Subcloning of the *dmyc* far Upstream Fragment: JD Construct

The *Drosophila* BAC clone RP98-34B12 contains the sequences of the *dmyc* far upstream region and is very similar to the BAC clone RP98-2A13. JD was obtained by polymerase chain reaction amplification of the RP98-34B12 BAC clone (obtained from the Children's Hospital Oakland Research Institute) in two steps to create products D1 and D2 (using a high fidelity polymerase chain reaction kit, Finnzymes Inc, Lafayette, CO). The amplified fragment D1 was 5588 bp in size, and the amplifying primers had the names PD1-F and PD1-R (Table 5.1). The amplified fragment D2 was 3440 bp in size, and the amplifying primers had the names PD2-F and PD2-R (Table 5.1). Polymerase chain reaction conditions in the thermal cycler were one cycle of initial denaturation at 98°C for 30 seconds; 30 cycles of

denaturation at 98℃ for 10 seconds, annealing at 56℃ for 30 seconds, and extension at 72℃ for 2 minutes; and one final extension cycle at 72℃ for 5 minutes, held at 4℃. The two polymerase chain reaction fragments were combined by blunt ligation (CIAP, Roche Diagnostics, Indianapolis, IN) into BlueScript to create SKII-D, then removed and exchanged for the J2.1 insert by 5' NotI and 3' Acc65I digestion:

JD = J2.1 X NotI X Acc65I + Insert SKII-D X NotI X Acc65I (8-kb band).

Fragment D1 was sequenced with the primers SEQD1-F1 to SEQD1-F9 (Table 5.1), Fragment D2 in JD was sequenced with the primers JD-F7 to JD-F9 (Table 5.1), and the plasmid JD was sequenced at the subcloning junctions using the primers pZR and pcaF (Table 5.1; No. 8 and No. 9).

### 4.4.3 Creation of the *dmyc* Intragenic Constructs

Plasmid pC-RP27 was digested with BamHI to isolate a 8943 bp 5' BamHI-BamHI 3' fragment containing *dmyc* intron2. The vector pBS-SKII+ was digested with BamHI at its MCS, dephosphorylated with CIAP (Roche Diagnostics). The vector and insert were ligated over-night to create plasmid SKII-In2fl. In this construct, the *dmyc* intron 2 fragment contains 3' ORF of Exon 2 at its 5' and 5' ORF of Exon 3 at its 3' end. The plasmid pC-RP27 served as template to mutate intron 2 fragment by PCR and remove the 5' and 3' end ORF sequences and to introduce a NotI restriction site at the 5' end an Asp718 restriction site at the 3' end of the fragment. The amplifying primers had the following names: In2-5F, In2-5R, In2-3F and In2-3R (Table 5.1; No. 19, 20 No., No. 21, and No. 22).

The PCR conditions for removal of 5' ORF were 1x initial denaturing at 94℃ for 3 min.; 30x denaturing at 94℃ for 20 sec., annealing at 57.8℃ for 30 sec., and elongation at 72℃ for 6 min; 1x final elongation at 72℃ for 8 min; ended at 4℃.

The PCR conditions for removal of 3' ORF were 1x initial denaturing at 94℃ for 3 min.; 30x denaturing at 94℃ for 20 sec., annealing at 55.3℃ for 30 sec., and elongation at 72℃ for 3 min; 1x final elongation at 72℃ for 8 min; ended at 4℃.

The 2654 bp In2.3'minus open reading frame PCR fragment and pBS-SK+ were digested separately with Asp718 to isolate the insert and vector bands, which were then ligated to get SKII-In2.3'-ORF intermediate template. The 2420 bp In2.5'minus

open reading frame PCR fragment and pBS-SK+ were digested separately with NotI to isolate the insert and vector bands, which were then ligated to get SKII-In2.5'-ORF intermediate template. The plasmids SKII-In2fl and SKII-In2.5'-ORF were digested separately with NotI and SgrAI to get the vector and the insert bands, which were then ligated into SKII-In2-ORF5. With this step 418 bp ORF fragment at the 5' end of the intron 2 sequence in SKII-In2fl was removed. The plasmids SKII-In2-ORF5 and SKII-In2.3'-ORF were digested separately with PmeI and Asp718 to obtain the vector and the insert bands, which were then ligated into SKII-In2-ORFs. With this step 371 bp ORF fragment at the 3' end of intron 2 was removed.

The construct SKII-In2-ORFs and the deletion construct J2.1 were digested separately with NotI and Asp718 to obtain the insert and the vector sequences. The insert and the vector DNA were ligated over-night at 16°C to get the J8 construct. With this step the ORF free intron 2 fragment was replaced with *dmyc* 5' fragment in J2.1 deletion construct.

For the creation of the J8.1 plasmid, the intermediate construct SKII-In2-ORFs was digested with NotI and PspOMI restriction enzymes to get the insert band. The pC-NLSlacZ was digested with NotI, dephosphorylated with CIAP (Roche Diagnostics) to obtain the vector band. The insert and the vector were ligated over-night at 16°C into J8.1. The NotI and PspOMI 5'-overhangs are compatible and ligate with each other (New England Bio Labs). The construct J8.1 is very similar to J8 except the orientation of the intron 2 fragment is 3'→ 5':

J8 = pC-NLSlacZ x NotI x Asp718 + Insert: SKII-In2-ORFs x NotI x Asp718 (8100 bp band)

J8.1 = pC-NLSlacZ x NotI+ Insert: SKII-In2-ORFs x NotI x PspOMI (8100 bp band).

The SKII-In2fl, SKII-In2.5'-ORF, SKII-In2.3'-ORF and SKII-In2-ORFs were sequenced with universal T3 and T7 primers at the subcloning junctions. The J8 and J8.1 were sequenced with designed sequencing primers pZR and pcaF (Table 5.1; No. 8 and No. 9) at the subcloning junctions.

The promoter in the transgene J8.4 is approximately 1-kb in size, and contains the enhancer of the large intron, just upstream of the DPE element (Figure 3.2). For the creation of J8.4, pC-RP27 served as the template, the amplified fragment was 2601

bp in size, and the amplifying primers had the names In2E-F-Not and In2E-R-Asp (Table 5.1, No. 42 and No. 43). For the creation of J8.5 insert, the DPE element of the large intron, J8 served as the template, the amplified fragment was 2662 bp in size, and the amplifying primers had the names pDPE-F and pDPE -R (Table 5.1, No. 44 and No. 45). Polymerase chain reaction conditions in the thermal cycler were one cycle of initial denaturation at 98℃ for 30 seconds; 30 cycles of denaturation at 98℃ for 10 seconds, annealing at 62℃ for 30 seconds, and extension at 72℃ for 2 minutes; and one final extension cycle at 72℃ for 5 minutes, ended at 4℃. The upper primer introduced a NotI site at the 5' end and the lower primer an Acc65I restriction site at the 3' end of the amplified product of the J8.4 promoter. The J8.4 polymerase chain reaction fragment was digested with NotI/Acc65I, and run on 1.5% agarose gel to isolate the 0.93-kb insert derived from the *dmyc* large intron fragment. The upper primer introduced a NotI site at the 5' end of the amplified product of the J8.5 promoter. The insert in the transgene J8.5 is approximately 2-kb in size, and is derived from 3' end of the *dmyc* large intron fragment that contains an Acc65I site at its 3' end. The polymerase chain reaction fragment was digested with NotI/Acc65I, and run on 0.5% agarose gel to isolate the 2-kb J8.5 promoter fragment. The isolated fragments of each construct was separately subcloned in 5' NotI-Acc65I 3' linearized, dephosphorylated (CIAP, Roche Diagnostics, Indianapolis, IN) pCaspeR4-NLSlacZ and pattB-temp del LoxP vectors in front of reporter *lacZ* to obtain the transgenes J8.4 and J8.5:

J8.4 = pC-NLSlacZ x NotI x Asp718 + Insert: pC-RP27 x NotI x Asp718 (1-kb band)

J8.5 = pC-NLSlacZ x NotI x Asp718 + Insert: J8 x NotI x Asp718 (2-kb band).

The insert of J8.5 was sequenced with designed sequencing primers Intr2-R 01, Intr2-R 02, Intr2-R 03, and Intr2-R 04 (Table 5.1; No. 50-53). Using the primers of J8 sequencing, pZR and pcaF (Table 5.1; No. 8 and No. 9), the constructs J8.4 and J8.5 were sequenced at the subcloning junctions. The oligonucleotide pcaF sequenced the J8.4 insert.

### 4.4.4 Fusion of 5'-UTR Truncation to the Intron2 Fragment in J9 and J10 Transgenes

J9 contains 2609 bp from the 3' end of the *dmyc* 5' promoter fused to intron 2 sequence minus ORFs to activate the expression of *lacZ* reporter. J10 is very similar to J9 except the orientation of the3' end of the *dmyc* 5' promoter is 3'→ 5'. The construct J2.1 was used as template to amplify the 2609 bp *dmyc* 5' promoter with PCR primers p9-F2 and p9-R2 (Table 5.1; No. 23 and No. 24). Primer 9-F2 introduces a PspOMI site at the 5' end and p9-R2 primer introduces a NotI site at the 3' end of the fragment. The PCR fragment and the J8 plasmid were digested separately with PspOMI/NotI to get the insert and vector bands. The vector and insert bands were ligated over-night at 16℃ into J9 and J10:

J9 = J8 x NotI + Insert: SKII-J5-promoter x NotI (2600 bp)

J10 = J8 x NotI + Insert: SKII-J5-promoter x NotI (2600 bp).

The subcloning junctions in J9 and J10 were sequenced with the sequencing primers pcaF and P9-R (Table 5.1; No. 9 and No. 25).

### 4.4.5 Generation of the *dmyc* 3'-End Truncations

To obtain the largest construct J1, a 5' BamHI-XbaI 3' ~10.5-kb fragment from the 3' end of *dmyc* was excised from for the plasmid pC-RP27, and replaced with the SV40 poly (A) in the construct J2.1. In the construct J1 the *dmyc* 3' fragment was used as trailer. Furthermore, this construct was used to build J1.1 *dmyc* 3' deletion construct.

The J1 construct contains the three predicted possible poly (A) signals of the *dmyc* gene (see Discussion, section 3.1.5).

J1.1 was created by deleting a 4759 bp 5' PshA1-Stu I 3' fragment and a 5943 bp 5' BglII-StuI 3' fragment, respectively from the 3' end of the ~10.5-kb *dmyc* trailer contained in J1. The linearized vector 5' PshA1-Stu I 3' was self-ligated to create J1.1, containing the poly (A1) signal. The junction was sequenced with sequencing primer pcaR (Table 5.1, No. 10).

In J1.3 all of the three poly (A) signals are removed but the sequences downstream of the predicted poly (A3) signal at the *dmyc* 3' are restored (Figure 2.4.2.d). A 2550 bp

fragment downstream of poly (A3) was amplified with primers pA0-F and pA0-R (Table 5.1, No 11 and No. 12); the primer pA0-R introduced a StueI site at the 3' end of the fragment. The PCR fragment and the vector J1 were each separately digested with AvrII/StuI; the insert and the vector were ligated to create J1.3. The 5' subcloning junction was sequenced with primer P1.3-F (Table 5.1, No. 26).

For the generation of the plasmid pJ.4, containing poly (A2), the 384 bp poly (A2) signal was amplified with the primers pA2-F-04 and pA2-R-05 (Table 1, No. 13 and 14). Primer pA2-F-04 introduced a BamHI site at the 5'-end of the fragment and the primer pA2-R-05 introduced an XbaI site at the 3'-end of the PCR fragment. The poly (A2) PCR fragment was blunt ligated into the vector pBS-SK+ (AleI linearized at its MCS) to obtain the intermediate substrate SK-poly (A2). The plasmid pJ5 and the SK-poly (A2) intermediate vector were digested separately with BamHI/XbaI to get the insert and the vector bands. The insert and the vector bands were ligated over-night at 16℃ to create J1.4. J1.4 was sequenced with sequencing primer pcaR (Table 1, No. 10).

To create J1.5, which contains poly (A3), the 440 bp poly (A3) fragment at 3'-end of J1 trailer was amplified with primers pA3-F and pA3-R (Table 1, No. 15 and No. 16). Primer pA3-F introduced a BamHI site at the 5' and pA3-R an XbaI site at the 3'-end of the fragment. The plasmid J5 and PCR fragment were digested separately with BamHI-XbaI to get the vector and insert DNAs. The vector and insert were ligated over-night at 16℃ to create J1.5. J1.5 was sequenced with sequencing primer pcaR (Table 1, No. 10):

J1 = J2.1 X BamHI X XbaI + Insert: pC-RP27 X BamHI X XbaI (10.5-kb band)

J1.1 = J1 (28420 bp) x PshA1 x StuI → vector band (23660 bp), self-ligation

J1.3 = J1 x AvrII x StuI + Insert: poly (A0)-PCR x AvrII x StuI (2550 bp band)

J1.4 = J5 x BamHI x XbaI + Insert: SK-Poly (A2) x BamHI x XbaI (384 bp band)

J.5 = J5 x BamHI x XbaI + Insert: poly (A2)-PCR x BamHI x XbaI (441 bp band).

### 4.4.6 Coordinates of Restriction Enzymes

a) *dmyc* 3' fragment: The coordinates for restriction sites in the genomic scaffold AE003427.1 (Gene Bank) are as follows: BamHI 142628 and XbaI 152590. BamHI begins at nucleotide -554 upstream of the translation stop and XbaI at nucleotide +9763 downstream of the translation stop.

b) *dmyc* 5' fragment: The coordinates of restriction sites in the genomic scaffold AE003427.1 are as follows: BamHI 5' 125738 and BamHI 3' 133687. BamHI 5' begins at nucleotide -7955 upstream of the translation start and BamHI 3' begins at nucleotide 803 downstream of the translation start.

c) Deletion fragments: NotI 7151 bp upstream of the *dmyc* translation start; XbaI 7142 bp upstream of translation start (both are in the polylinker of the plasmid SKII-dmyc5); MscI 5490 bp upstream of the translation start; NheI 4822 bp upstream of the translation start; SpeI 4125 bp upstream of the translation start; NarI 2578 bp upstream of the translation start; SphI 1913 bp upstream of the translation start; Asp718 was introduced just before the translation start.

d) RP27 fragment: NotI 5424 bp upstream of the *dmyc* gene start site; DraIII 5817 bp upstream of the *dmyc* gene start site (both are in the polylinker of the BAC clone RP98-2A13); XbaI 21885 bp downstream of the *dmyc* start site.

### 4.4.7 Plasmid DNA Preparation for Injection

The Electro Max DH10B *E. coli* strain was electrotransformed each time with one of the plasmids J1-J7,J1, J1.1, J1.3-J1.5, J8, J8.1 J8.2, J8.4, J8.5, J9, J10, and all the intermediate constructs that served as substrate for obtaining the above constructs. Single isolated colonies were grown on 2x LB-ampicillin plates over-night.

Isolated colonies were analyzed according to the rapid alkaline lysis miniprep method (Children's Hospital Oackland Research Institute CHORI, U.S.A). 5 µl of the miniprep DNA was checked on a 0.6% agarose gel for the identification of the right colony.

For the injection plasmid DNA J1, J1.1, J9, and J10 Large Construct Kit (Qiagen) was used. For the plasmid DNA from the rest of the transgenes, QIAfilter plasmid Midi Kit was used (Qiagen).

### 4.4.8 Chromosomal Linkage Analyses of the Inserted P-Elements

Embryos of the genotype $y^1$ $w^{1118}$ were injected for random P-element transformation. In order to determine the chromosome linkage and balance the inserted P-element, for all the constructs the following crosses were set up:

- The single eclosed G0 flies were crossed to the double balancer line: *w*; Cyo; *TM3, Sb Ser* (Bloomington Stock Number 2475) (Table 2.3, No. 10).

- The transgenic F1 flies were crossed to $y^1$ $w^{1118}$ flies to establish lines.

- For the insertions on the X chromosome the balancer line *FM7c* was used. For the insertions on the second chromosome the *CyO* line was used, and for the insertions on the third chromosome the balancer *TM3* was used.

The larvae, ovaries, and the embryos for testing *lacZ* activity were taken from F2 flies and the later generations.

## 4.5 β-Gal Detection by X-Gal Reaction

For each construct, 4–15 independent transgenic lines (except the largest construct [J1] for which only two independent lines were obtained) were dissected and X-Gal staining was performed at standardized reaction conditions (Song et al., 2004). The incubation temperature for different tissues was as follows: discs 29°C, embryos 37°C, and ovaries at room temperature. For each reaction, *dpp-lacZ* fly stocks (gifted by Dragan Gligorov, Karch's laboratory) were used as a positive control, while *y[1] w[1118]* and attP-fly stocks were used as negative controls.

### 4.5.1 Reverse Transcription Polymerase Chain Reaction Analysis

Total RNA was isolated from the sample discs and ovary (Figure 2.5.2.b, Figure 2.5.3.c, and Figure2.5.4.d) using an aMResco phenol-free total RNA purification kit (Code N788 kit) with RNase-free DNase treatment (Promega, Basel, Switzerland) following the manufacturer's protocol. Total RNA 1 µg was reverse transcribed into cDNA in a reaction volume of 30 µL using SuperScript™ III reverse transcriptase oligo(dT) 20 primers and reverse transcription reagents from Invitrogen (Carlsbad,

CA). Semiquantitative polymerase chain reactions were performed on the resulting cDNA using a high fidelity Phusion DNA polymerase kit (Finnzymes, Bioconcept, Switzerland). Polymerase chain reaction primers were designed for the *lacZ* reporter gene and *Drosophila* actin gene using PrimerSelect from the Lasergene software suite (DNASTAR, Madison, WI). All primers were synthesized at Microsynth. Primer sequences are indicated from 5' to 3' in Table 1 (primers No. 46-49). The polymerase chain reaction conditions in the thermal cycler were one cycle of initial denaturation at 98℃ for 20 seconds; 30 cycles denaturation at 98℃ for 10 seconds, annealing at 61.3℃ for 30 seconds, and extension at 72℃ for 50 seconds; one cycle final extension at 72℃ for 2 minutes, held at 4℃. The polymerase chain reactions, 5 µL per lane, were run on 1% agarose gel for 90 minutes at a voltage of 120.

### 4.5.2 Detection of lacZ Protein by Immunoblotting

For Western Blot analysis, ovaries taken from the flies carrying the J8 or J8.5 transgene (Figure 2.5.4.e) were treated with Cytobuster™ Protein Extraction Reagent (71009-3, Novagen). The extracts were treated with phreon to remove the yolk proteins. Proteins extracted were run on SDS-PAGE denaturing system (20 µg/lane) using PowerEase® 500 Power Supply (invitrogen) for 40 minutes at a constant voltage of 200V. Proteins were then transferred to nitrocellulose membrane using iBlot® Western Detection Kit (Life technologies, IB7410-01). The membranes were processed using WesternBreeze® Chromogenic Kit–Anti-Rabbit (WB7105, invitrogen). lacZ protein was visualized with ANTI-BETA-GALACTOSIDASE, (Molecular Probes, A11132) and *Drosophila* actin was detected with Anti-Actin monoclonal antibody (Millipore, MAB150IR).

## 4.6 Rapid Amplification of the *dmyc* cDNA Ends (5' RACE)

RACE Ready cDNA was prepared using a Smarter™ Race cDNA Amplification Kit (Cat. No. 634924, Clontech Laboratories, Inc.), and following the manufacturer's user manual (Protocol No. PT4096-1, Version No. 011312). For the synthesis of cDNA for the positive control reactions, mouse heart total RNA (1µg/µl) and for the generation of *dmyc* cDNA, adult *Drosophila* total poly($A^+$) RNA (1µg/µl) (Clontech Laboratories, Inc.) were used. Total RNA, 1 µg, was reverse transcribed into cDNA in a reaction

volume of 10 µL using SMARTScribe reverse transcriptase and SMARTer IIA oligo. Polymerase chain reactions (PCR) of the *dmyc* and the positive control cDNAs were performed using Advantage® 2 Polymerase Mix and the Universal Primer Mix (UPM) (Cat. No. 639201, Clontech Laboratories, Inc.). RACE PCR reactions were performed with UPM forward primer (recognizes the SMARTer sequence at the 5'-end of the cDNA) and the gene specific primers GSP-N01, GSP-C05, and GSP-06 reverse primes (Table 5.1, No. 54-56). Thermal cycler for PCR 1 (Figure 2.6.a) was commenced using the following program for touchdown PCR: five cycles of initial denaturation at 94°C for 30 seconds and denaturation at 72°C for 3 minutes; 5 cycles of denaturation at 94°C for 30 seconds, annealing at 70°C for 30 seconds, and extension at 72°C for 3 minutes; and 25 cycles of denaturation at 94°C for 30 seconds, annealing at 70°C for 30 seconds, and extension at 72°C for 3 minutes. For the PCR 2 (Figure 2.6.a) the stringency of the reactions in the thermal cycler was increased by raising the annealing temperature in increments of 2°C.

The amplification products of both PCR reactions were examined by running on 1.5% agarose gel. The product of N01 was sequenced with the primers Exn1-R2 and Exn1-R3 (Table 1, No. 59 and No. 60). The products of C05 and C06 were sequenced with the primers Exn2-R1 and Exn2-R2 (Table 1, No. 61 and No. 62).

# 5. Appendix

## 5.1 List of the Oligonucleotides Designed for and Used the Study

All BAC clones, intermediate plasmid constructs, engineered transforming vectors, and transgenes were sequenced by Microsynth AG, Balgach, Switzerland. Except for the standard primers (provided by Microsynth), all the other oligonucleotides and sequencing primers were designed using the software tool DNASTAR Lasergene 9.1 Module Primer Select and synthesized at Microsynth AG. Sequences for all the polymerase chain reaction, reverse transcription polymerase chain reaction, 5' RACE, and sequencing primers are listed in Table 5.1.

**Table 5.1.a** Analytical primer pairs

| No | Name | Primer Sequence (5'→3') | bp | Usage |
|---|---|---|---|---|
| 1 | PX1 | ACG GTT GGC ATT GTT GAC TC | 20 | BAC clone RP98-2A13 sequencing |
| 2 | PN1 | AAT TCG TTT TCC ATA GAT A | 19 | |
| 3 | BAC-F | CTT CGC GTC CAA CAG ATG | 18 | pC-RP27 sequencing |
| 4 | BAC-R | GCG CCA AAG CAA GAG GGA ATG | 21 | |
| 5 | pZpA-F | CCA AAA AAG GGT TTC ATT AAC TTG TAC ACA TAC | 33 | pNLS-LacZ sequencing |
| 6 | pZpA-R | TTG GGG ATT TTA ATA GCG GGC CCT GTG TGT | 30 | |
| 7 | SV40 | GTT CAG GGG GAG GTG TGG G | 19 | SV40 P(A) sequencing |
| 8 | pZR | CGG GCC TCT TCG CTA TTA CG | 20 | pJ8 sequencing |
| 9 | pcaF | GAC GGC GAT ATT TCT GTG GAC | 21 | pCaSpeR4 and J8.4 sequencing |
| 10 | pcaR | CCT TAG CAT GTC CGT GGG GTT TGA | 24 | |
| 11 | P(A0)-F | CCC CCT TGC TAT AAC CCG TAT AT | 23 | PCR – pJ1.3 |
| 12 | P(A0)-R | AAA CGG TTT CTA TAT AT AGGCCTG TGT GTG T StuI | 31 | |
| 13 | pA2-F-04 | TTT GCT GCC ACT TCT GTGGATCCT CAT ACA TT BamHI | 32 | PCR – pJ1.4 |
| 14 | pA2-R-05 | GGT TTT AGC GGC TTT GAT CGG TCTAGA AAT AAT ATT TT XbaI | 38 | |
| 15 | pA3-F | AAG GGG AAA CTG ATT GGATCC AAA AAG CAC A BamHI | 31 | PCR – pJ1.5 |
| 16 | pA3-R | CGC CTC GCT TTC ATA CAA TCTAGA TAA TGT TC XbaI | 32 | |
| 17 | dm5E2F | GCG CCA AAG CAA GAG GGA ATG AAC | 24 | PCR – pJ2.1 |
| 18 | dm5E2R | CGG GAG AG GGTACC TGC GAT TAT GTT GTC T GGG TTT TTT TTT C ACC651 | 43 | |
| 19 | In2-5F | GAC CCC CTGCGGCCGCGG TGA GTC AAA TTT ATA TAC TTT T NotI | 40 | PCR – In2.5' |
| 20 | In2-5R | CCC CGT CGCGGCCGC TCA CGT AAG CTT TTA TAC TAA T NotI | 37 | |
| 21 | In2-3F | GCA ACA GGTACC GGC TTG TGT GCA TTT TAT TT Acc65I | 32 | PCR – In2.3' |
| 22 | In2-3R | TCG ATT GGTACC TCT GCA CAG CGA TAG TAA AAA AA Acc65I | 35 | |

**Table 5.1.b** Analytical primer pairs

| No | Name | Primer Sequence (5'→3') | bp | Usage |
|---|---|---|---|---|
| 23 | P9-F2 | TGT AGGGCCCCA AAG CAA GAG GGA ATG AAC AA PspOMI | 32 | PCR pJ9 |
| 24 | P9-R2 | GTGCGGCCGCTG CGA TTA TGT TGT CT NotI | 32 | |
| 25 | P9-R | TTG TGT GTG AGG GCC CTG CGA TTA TGT TGT CT | 32 | pJ9 sequencing |
| 26 | P1-3-F | GTT GGA CGA CGC GAA GAT GAA AGA GAA | 27 | pJ1.3 sequencing |
| 27 | pJ8.1-F | TGA GCG CGC GTGGGCCCA CTC ACT ATA PspOMI | 27 | PCR pJ8.1 |
| 28 | pJ8.1-R | CATGTACAT GGCGGCCGC TCA AGA CAC C BsrGI NotI | 28 | |
| 29 | PD1-F | GGG CAC ATA ACA ACA TCT GAG GAA ACG AT | 29 | PCR D1 |
| 30 | PD1-R | AAG GGA GAG GGG CTA ACC ACT CAATAC G | 28 | |
| 31 | PD2-F | AAG GTT TTG GCC CAC CGT TTG AAA CAT ACA TAC | 33 | PCR D2 |
| 32 | PD2-R | ATT TGG GGG TAC CTT TCA CGG AGT ATC CTA TTT | 33 | |
| 33 | SEQD1-F1 | GCC TTT TCC GTG TAG CGT ATC GTA | 24 | Fragment D1 sequencing |
| 34 | SEQD1-F2 | TTC GAT TTT CCG CTG AAG CCA AGA CAG G | 28 | |
| 35 | SEQD1-F3 | AAG TGG TGC TTT GCT AAC CGC CCC CTG TT | 29 | |
| 36 | SEQD1-F4 | TGC TAA ACC CAC CAT CGC CTA TCT TG | 26 | |
| 37 | SEQD1-F5 | TGC GCC ATC GAT TGC CAG TTC TTA CAC | 27 | |
| 38 | SEQD1-F6 | ACG AGG TTC AGG GTC AGC GAT GTA AT | 26 | |
| 39 | SEQD1-F7 | GGG CCT GCC AAA ATC CAT AA | 20 | |
| 40 | SEQD1-F8 | GAA TGC CTA AAC CTC TTG TAA T | 22 | |
| 41 | SEQD1-F9 | ATT TTG TTT TTG GCT TGC TGT CTT ATT AG | 29 | |
| 42 | In2E-F-Not | GCA CTC TTA ACA A GCGGCCGC ATG TTC.TTT TT NotI | 32 | PCR – pJ8.4 |
| 43 | In2E-R-Asp | AAC A GGTACC AC CGC TGT GTT ATA AAC AATTGC G Acc65I | 34 | |
| 44 | pDPE-F | CCG ATT TGT GCGGCCGC AATAAC CGT TAG TAA AAC T NotI | 36 | PCR – pJ8.5 |
| 45 | pDPE-R | ATG CGC TCAGGTCAAATTCAGACG | 24 | |

**Table 5.1.c** Analytical primer pairs

| No | Name | Primer Sequence (5'→3') | bp | Usage |
|---|---|---|---|---|
| 46 | ACT-F01 | CGCTCCCCGTGCTGTCTTCC | 20 | RT-PCR: Actin Detection |
| 47 | ACT-R01 | GCGGTGGCCATCTCCTGCTC | 20 | |
| 48 | lacZ-pcr F | TGGTTATGCCGATCGCGTCACACT AC | 26 | RT-PCR: lacZ Detection |
| 49 | lacZ-pcr R | CGGATAAACGGAACTGGAAAAACT GC | 26 | |
| 50 | Intr2-R 01 | CCT CAT CTG CAC AGC GAT AGT A | 22 | pJ8.5 sequencing |
| 51 | Intr2-R 02 | TCA TCT GCA CAG CGA TAG TAA | 21 | |
| 52 | Intr2-R 03 | CTG CAC AGC GAT AGT AAA | 18 | |
| 53 | Intr2-R 04 | GGG GAA GGG GGA GGT GGA GAA GA | 23 | |
| 54 | GSP-N01 | GCC GGG CGG TAT TAA ATG GAC CTC GTC CTG | 30 | 5' RACE cDNA synthesis |
| 55 | GSP-C05 | CTA GCC CTA CGC CGC CGC TTT AAG CCG ATA GTA TGA GAC | 39 | 5' RACE cDNA synthesis |
| 56 | GSP-C06 | CAC TGC GCC GCT GCC AAT TAC GAT CCA TG | 29 | 5' RACE cDNA synthesis |
| 57 | UPM-sh | CTA ATA CGA CTC ACT ATA GGG C | 22 | 5' UTR RACE cDNA sequencing |
| 58 | Exn1-R2 | AAA CAG GCA AAA ATC CAA TCC ACA | 24 | 5' UTR RACE cDNA sequencing |
| 59 | Exn1-R3 | GGT TAT TTC TGG GAT TTA GCA CTT | 24 | |
| 60 | Exn2-F | ACC GGC CAG CAG CAG TCC A | 19 | 5' RACE cDNA sequencing |
| 61 | Exn2-R1 | GAA TGG TGG ACG ACG AAA GGA GTT | 24 | |
| 62 | Exn2-R2 | CTC TGG CTT CAT ATC CTC CTC TAA | 24 | |

**Table 5.1. All the oligonucleotides used in the study are listed.** The oligonucleotides were designed with Bioinformatics tool Laser Gene 9.1 module Primer Select and synthesized at Microsynth AG (Balgach, Switzerland). As indicated, the primer sets were used for sequencing and amplification of *dmyc* non-coding regions as well as for sequencing of cloning vectors and injection plasmids. Orientation of the indicated primer sequences is 5' → 3'. Length of primers is in base pairs (bp).
**Abbreviations:** bp, base pairs; RT, reverse transcriptase; PCR, polymerase chain reaction; E, exon; F, forward; R, reverse; RACE, rapid amplification of cDNA ends; Exn, exon; Intr, intron; DPE, downstream promoter element; UPM-sh, universal primer short.

## 5.2 Results of the Positive and Negative Controls

For each β-gal reaction in the Results sections, a positive and a negative control experiment was carried out under the same conditions. In all the experiments the fly stock "Blue Balancer" and the enhancer trap line $y^1$ $w^{1118}$; *dpp-lacZ* was used as a

source for positive control in larval brains and discs, in the embryos, and ovary staining (Table 2.3). For the negative controls, dissected brains and discs, embryos, and ovarian tissues from $y^1$ $w^{1118}$ and different attB fly lines were used (Table 2.3). The results of positive and negative staining reactions are presented in this section.

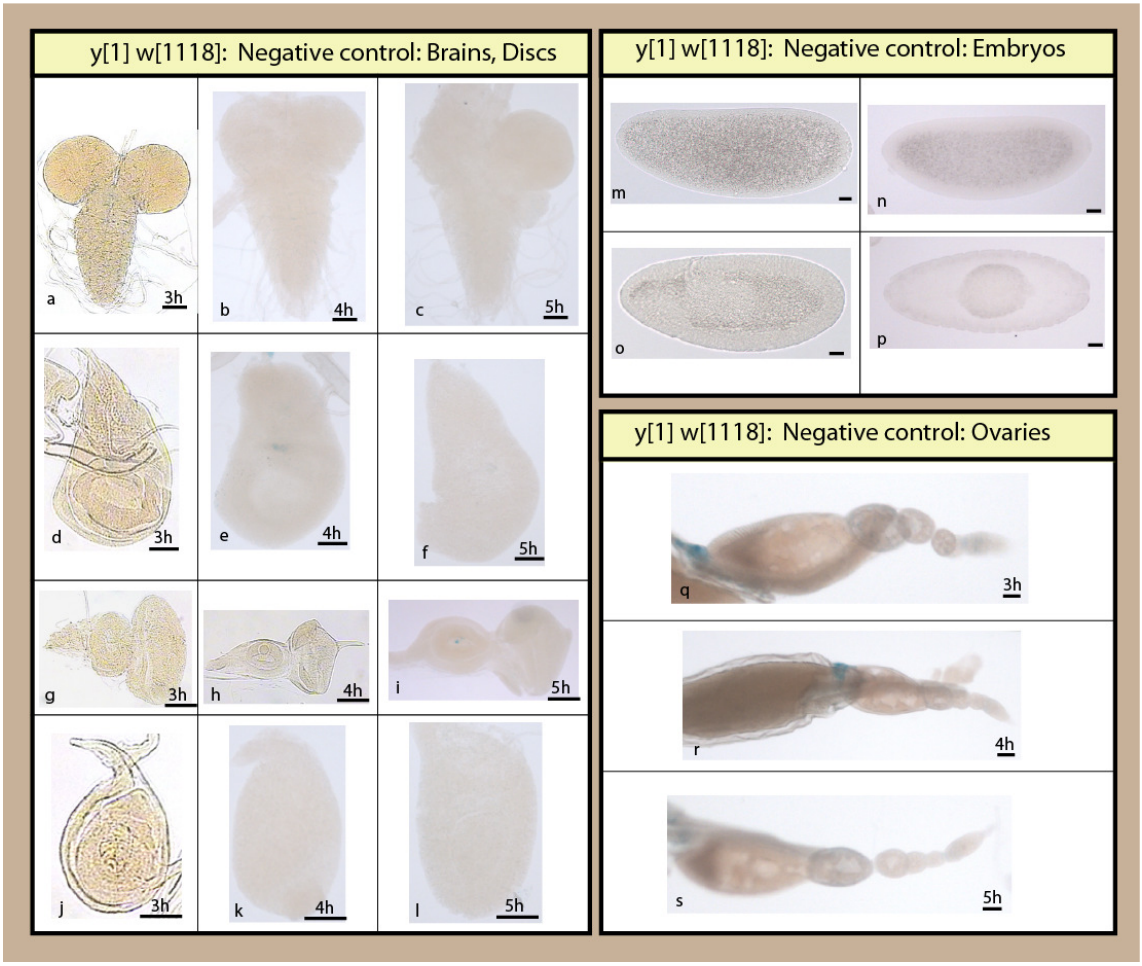



**Figure 5.2.a. Negative controls for 3, 4, and 5 hours lacZ staining.** For each lacZ staining of transgenes negative control staining was performed. As negative control different tissues were taken from the $y^1$ $w^{1118}$ flies listed in Table 2.3. Depicted are negative controls (**a-s**) (**a-c**, brain; **d-l**, discs; **m-p**, embryos; and **q-s**, ovaries). Staining times for the discs and ovaries are indicated above the scale bars. Embryo staining took place over-night. Scale bar in (**a-s**) indicates 50 µm.

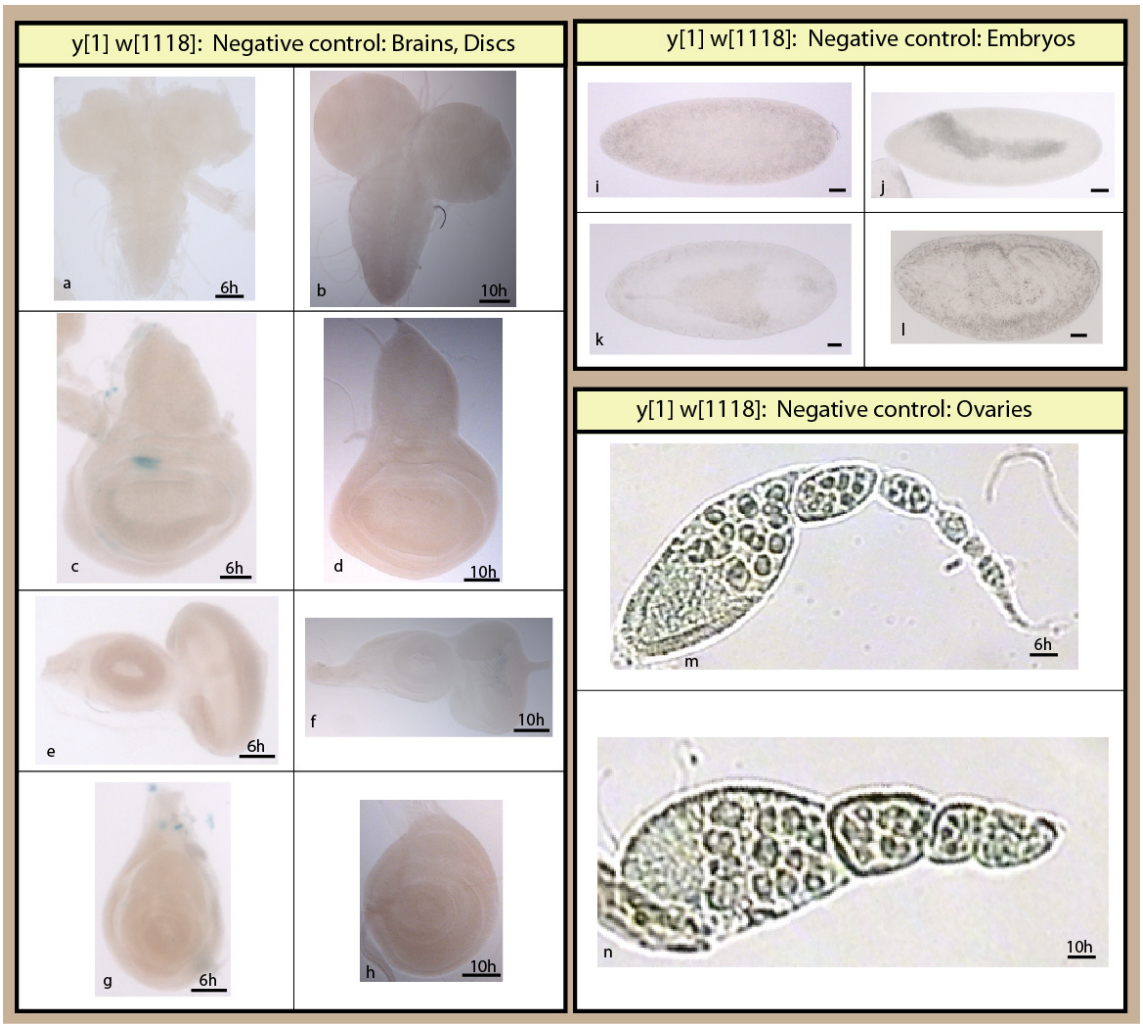



**Figure 5.2.b. Negative controls for 6 and 10 hours lacZ staining.** Different tissues dissected from $y^1$ $w^{1118}$ flies served as negative control for the staining with the tissues taken from the *dmyc-lacZ* transgenes. Depicted are negative controls (**a-n**) (**a**, **b**, brain; **c-h**, discs; **i-l**, embryos; and **m**, **n**, ovaries). Staining times for the discs and ovaries are indicated above the scale bars. Embryo staining took place over-night. Scale bar in (**a-n**) indicates 50 µm.

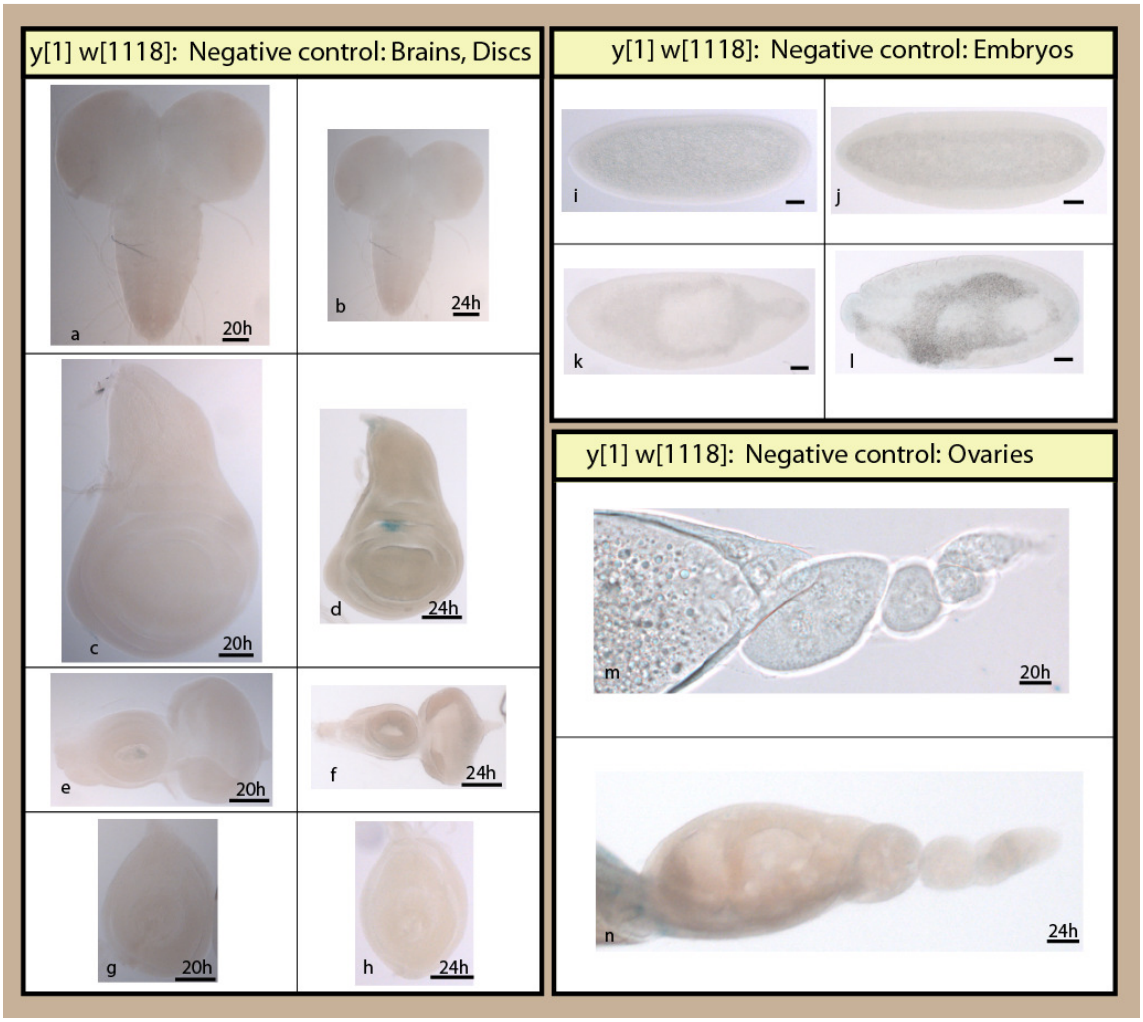



**Figure 5.2.c. Negative controls for 20 and 24 hours lacZ staining.** As indicated in Figure 5.3.a and Figure 8.3.b, tissues from $y^{1}$ $w^{1118}$ flies served as negative control for the staining with the tissues taken from the *dmyc-lacZ* transgenes. Depicted are negative controls (**a-n**) (**a**, **b**, brain; **c-h**, discs; **i-l**, embryos; and **m**, **n**, ovaries). Staining times for the discs and ovaries are indicated above the scale bars. Embryo staining took place over-night. Scale bar in (**a-n**) indicates 50 µm.

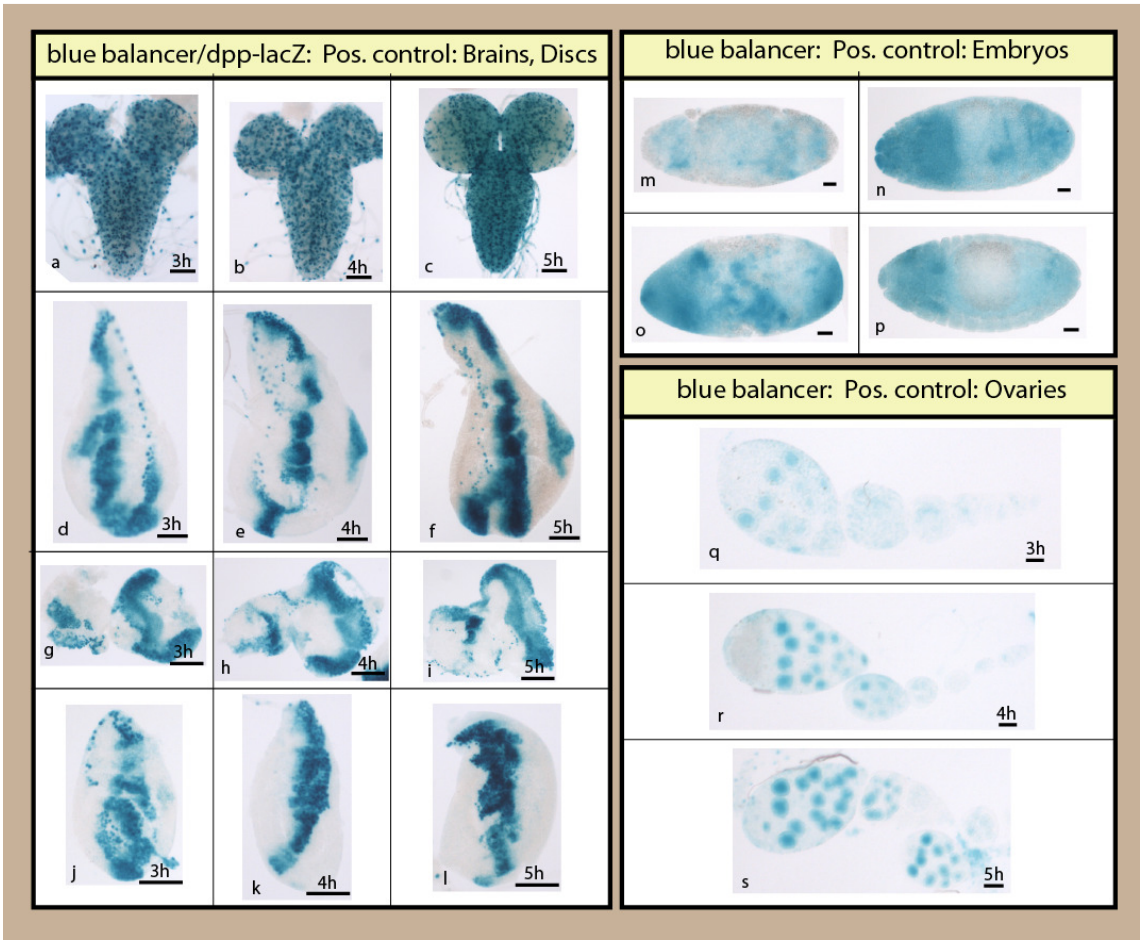



**Figure 5.2.d. Positive controls for 3, 4, and 5 hours lacZ staining.** For each lacZ staining of transgenes positive control staining was performed. Blue Balancer and *dpp*-lacZ flies served as source for positive control (Table 2.3). Depicted are positive controls (**a-s**) (**a-c**, brain: Blue balancer; **d-l**, discs: *dpp-lacZ*; **m-p**, embryos; and **q-s**, ovaries). Staining times for the discs and ovaries are indicated above the scale bars. Embryo staining took place over-night. Scale bar in (**a-s**) indicates 50 µm.

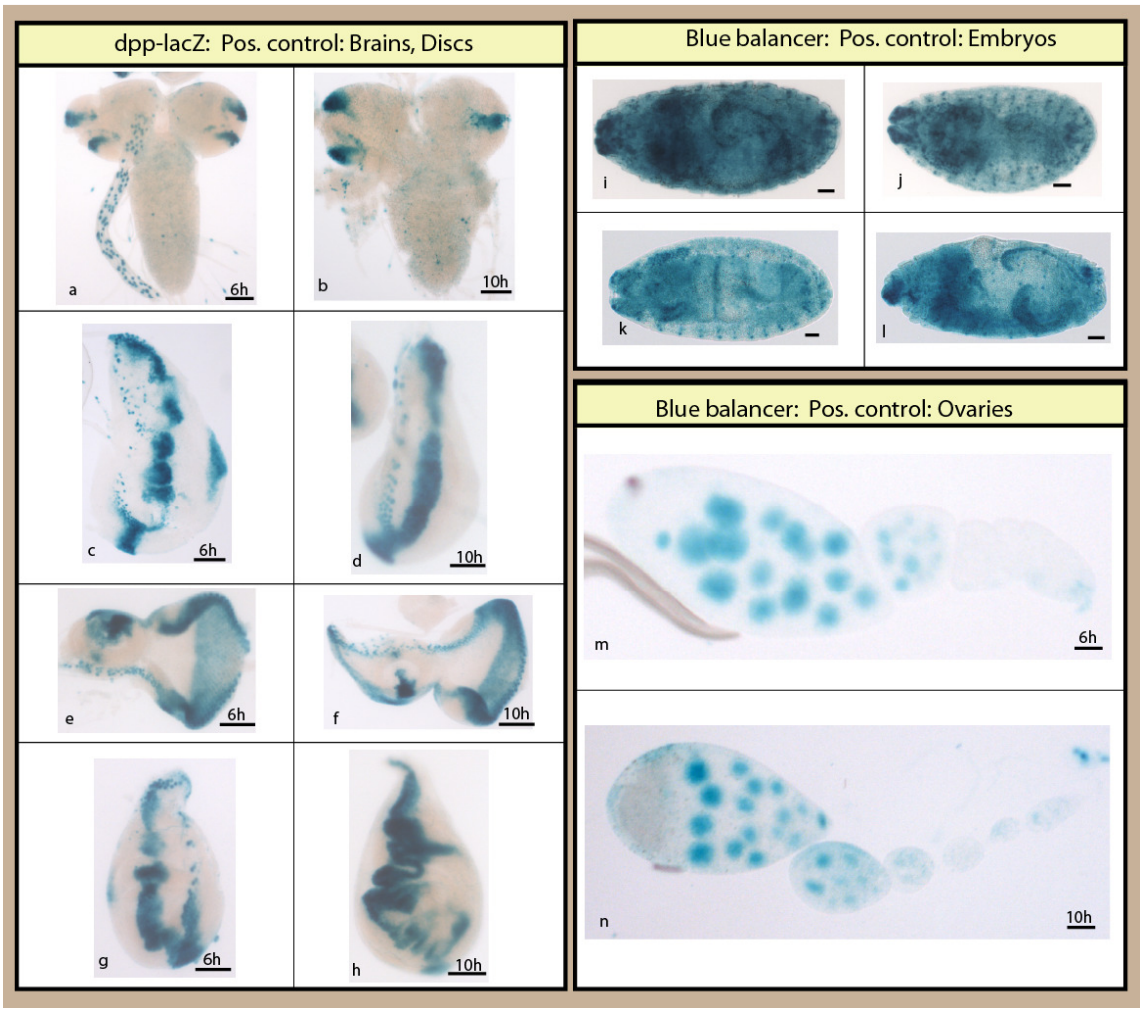



**Figure 5.2.e. Positive controls for 6 and 10 hours lacZ staining.** Dissected tissues from Blue Balancer and *dpp-lacZ* flies were used as positive control staining. Depicted are positive controls (**a-n**) (**a**, **b**, brain; **c-h**, discs; **i-l**, embryos; and **m**, **n**, ovaries). Staining times for the discs and ovaries are indicated above the scale bars. Embryo staining took place over-night. Scale bar in (**a-n**) indicates 50 µm.

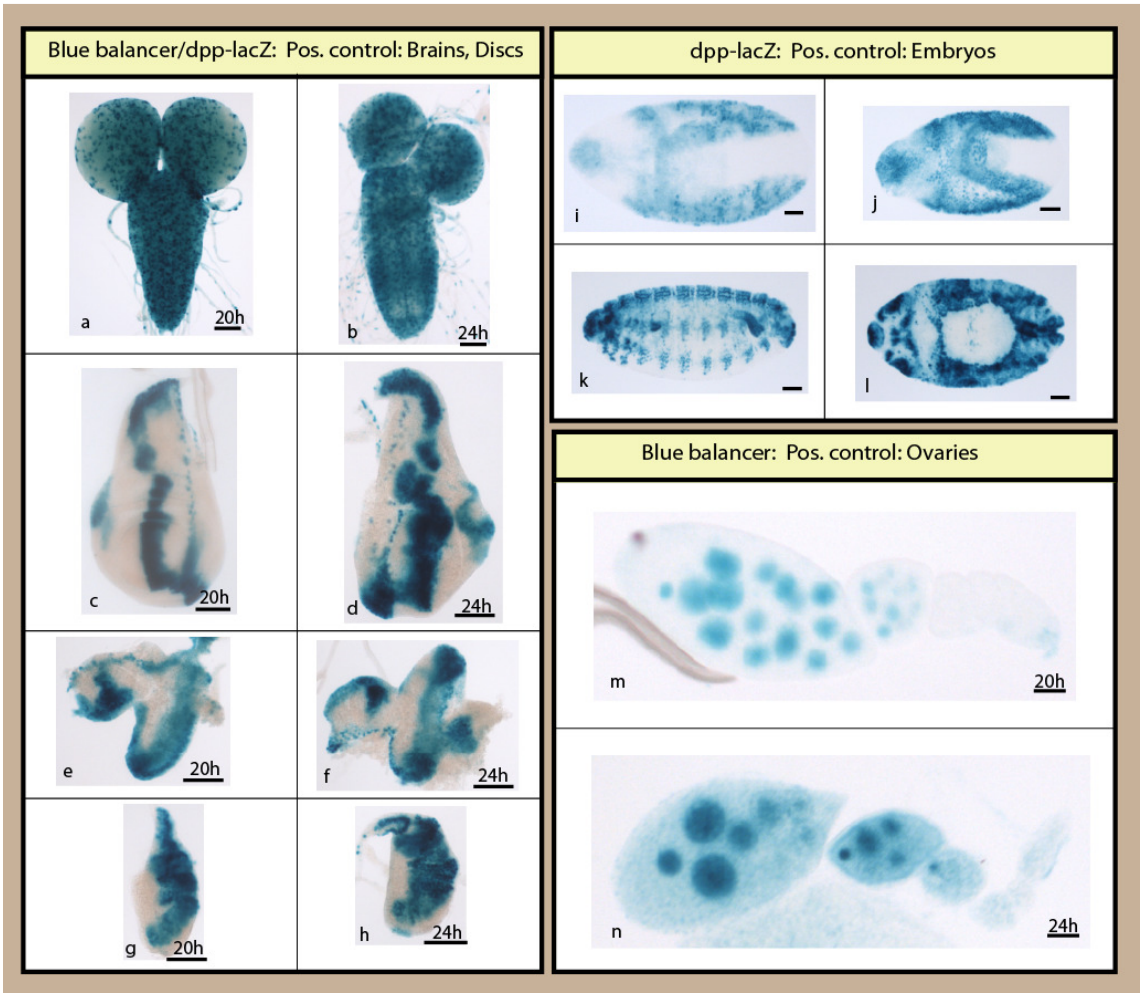



**Figure 5.2.f. Positive controls for 20 and 24 hours lacZ staining.** As indicated in Figure 5.2.d and Figure 5.2.e, for each lacZ staining of transgenes positive control staining was performed. Depicted are positive controls (**a-n**) (**a**, **b**, brain; **c-h**, discs; **i-l**, embryos; and **m**, **n**, ovaries). Staining times for the discs and ovaries are indicated above the scale bars. Embryo staining took place over-night. Scale bar in (**a-n**) indicates 50 µm.

## 5.3 Major Vector Maps Relevant to the Study

In the following section the restriction maps (see Section 2.4) of the main plasmids processed during this work as well as the map of the BAC clone RP98-2A13 will be shown.

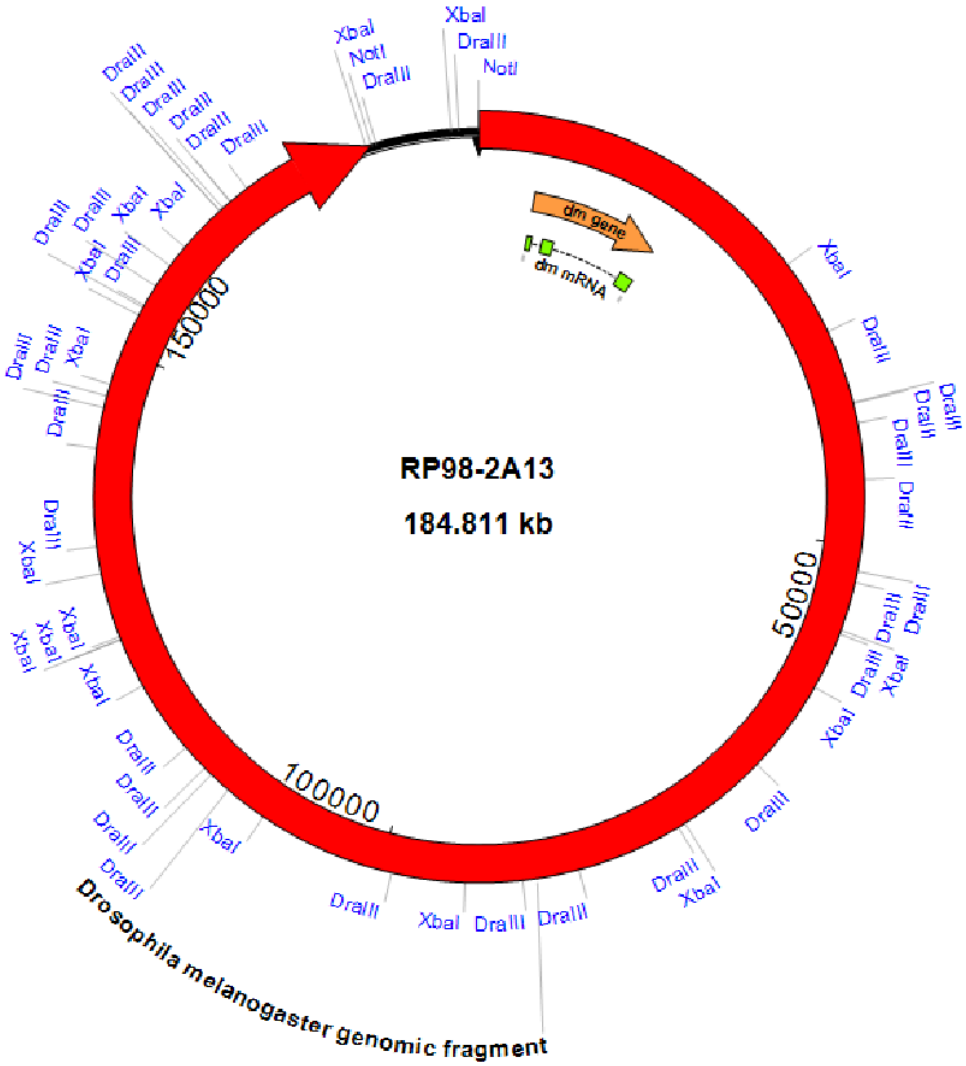



**Figure 5.3.a. Circular map of the BAC clone RP98-2A13.** RP98-2A13 consists of the BAC vector pBACe3.6 and 175996 bp of the *Drosophila* X-chromosome genomic sequence (red arrow). The genomic scaffold with the Gene Bank accession number AC108480 contains the *dmyc* gene (orange arrow).

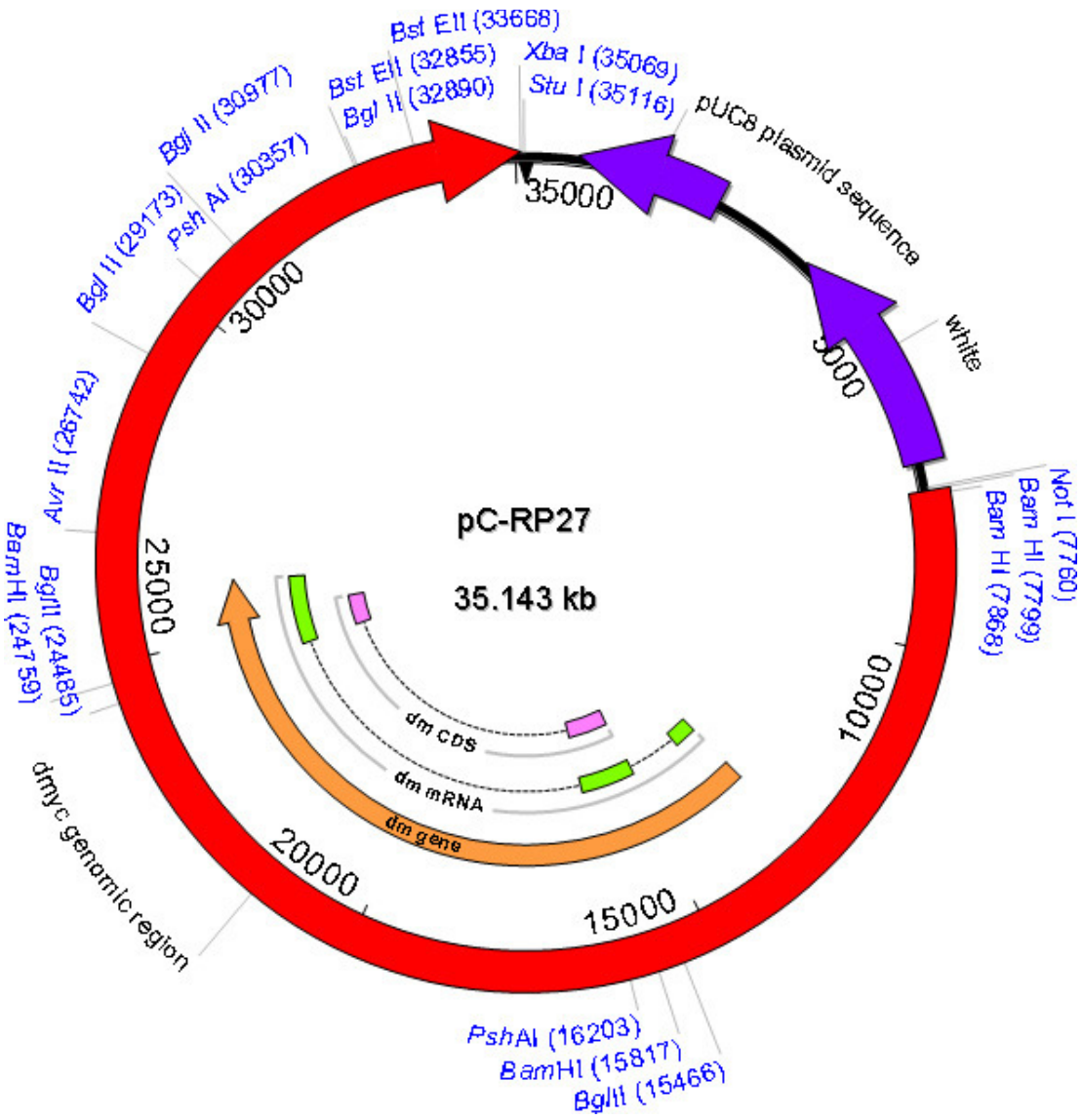



**Figure 5.3.b. Plasmid map of pC-RP27.** The map shows the ~27-kb 5'NotI-XbaI3' fragment containing the *dmyc* locus (red and orange arrows). The ~27-kb fragment was excised from the BAC clone RP98-2A13 by restriction digestion of the BAC clone with the enzymes DraIII, NotI and XbaI and subcloned into the fly transforming vector pCaSpeR4.

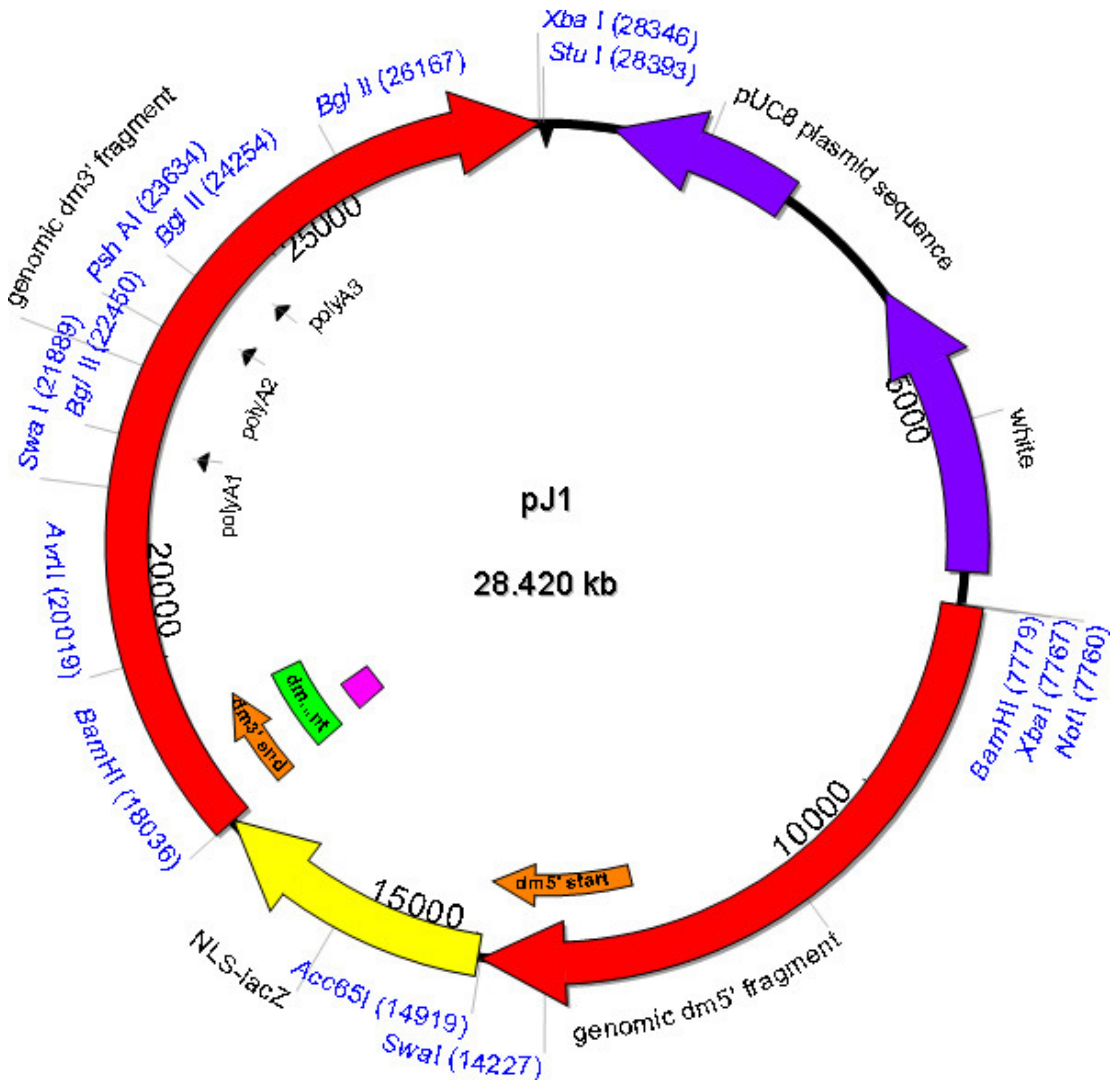



**Figure 5.3.c. J1 plasmid circular map.** This construct is the largest transgene created during the work. It contains the *dmyc* 5' and the *dmyc* 3' full-length fragments. The three predicted polyA1, polyA2 and polyA3 sites are indicated (black arrows). This plasmid was used as substrate to create the J1 deletion constructs (maps not shown).

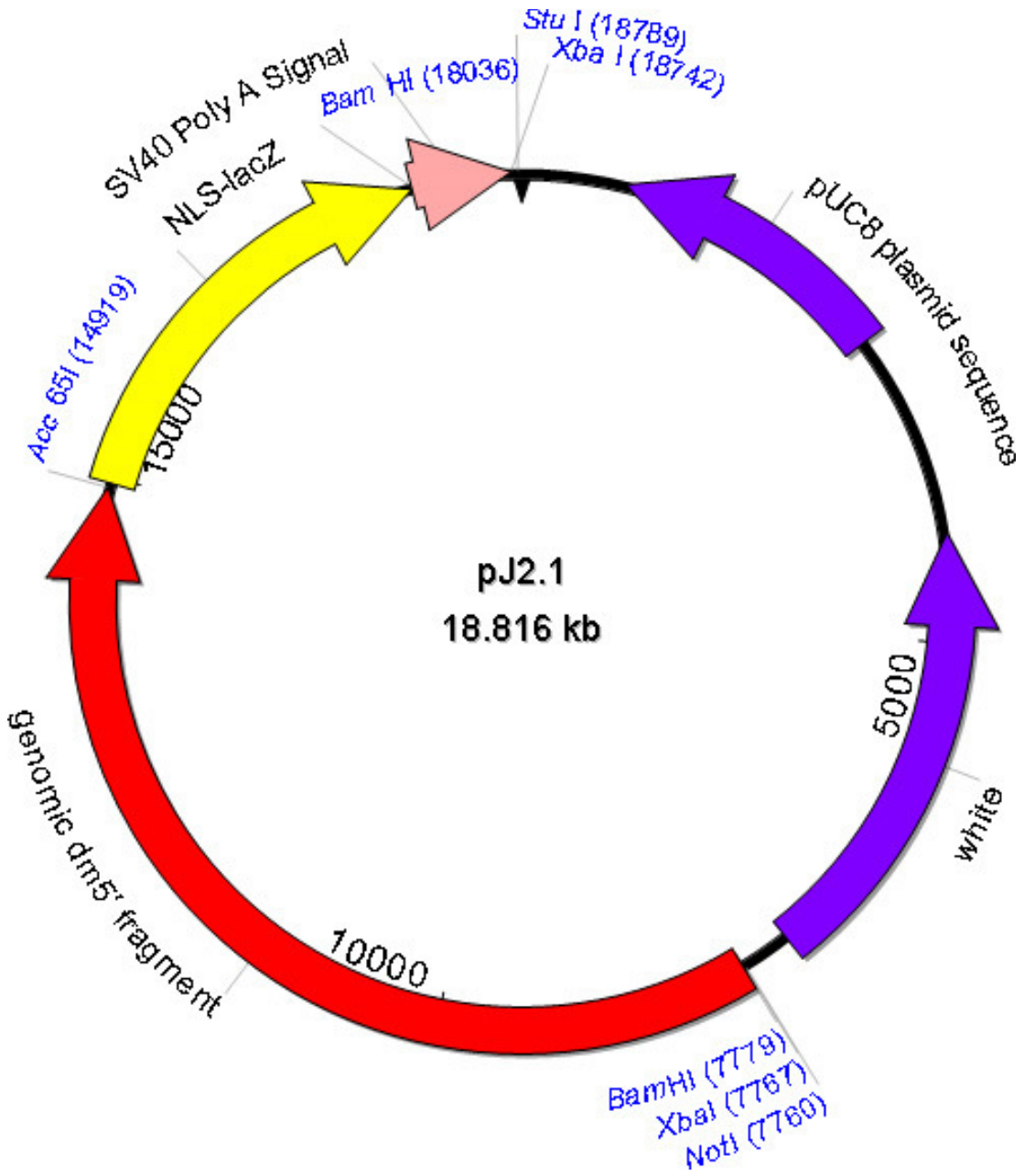



**Figure 5.3.d. J2.1 plasmid circular map.** This plasmid contains the *dmyc* 5' full-length fragment. As shown on the map, SV40 polyadenylation signal terminates the transcription of the reporter gene *NLS-lacZ*.

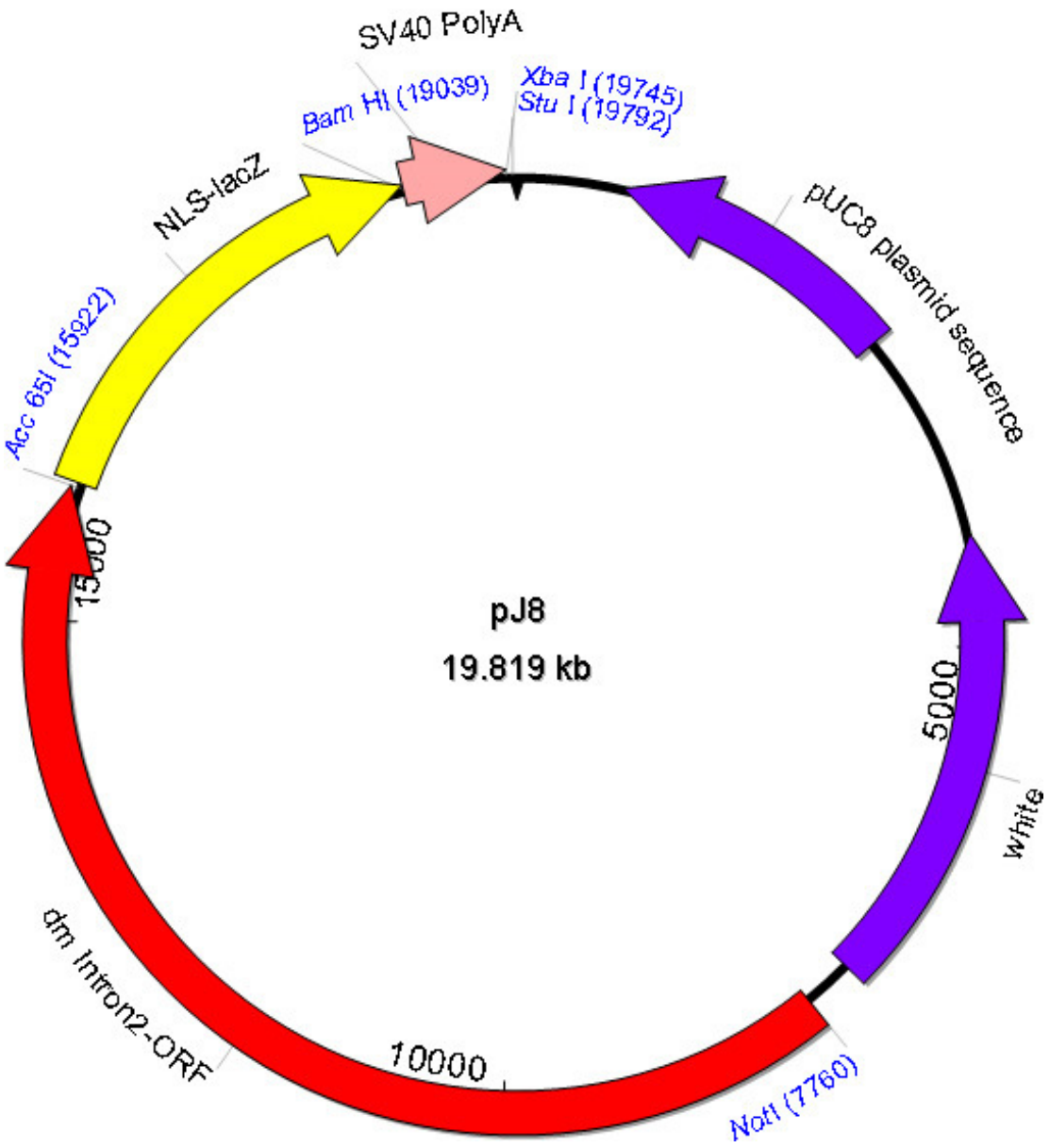



**Figure 5.3.e. Circular map of the pJ8 construct.** This construct contains the ~8.1-kb *dmyc* intron 2 fragment, which is free from the 5'- and the 3'-end coding sequences. The corrected (ORF free) intronic sequence is fused to the reporter *lacZ* and the transcription terminator SV40-polyA signal.

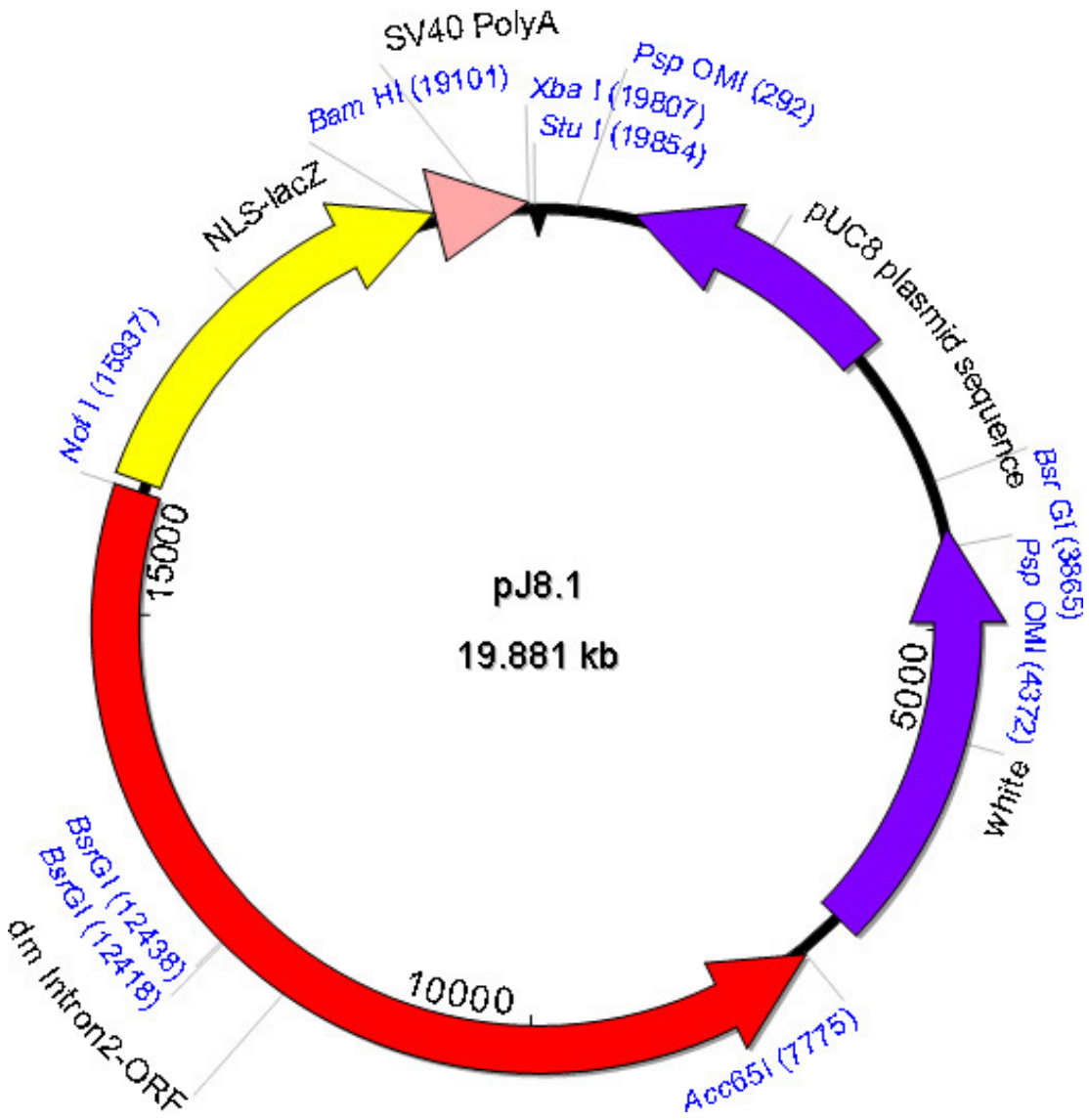



**Figure 5.3.f. Circular map of the pJ8.1 construct.** This construct is very similar to J8 except that the orientation of the *dmyc* intron 2 fragment is reverse (3' → 5').

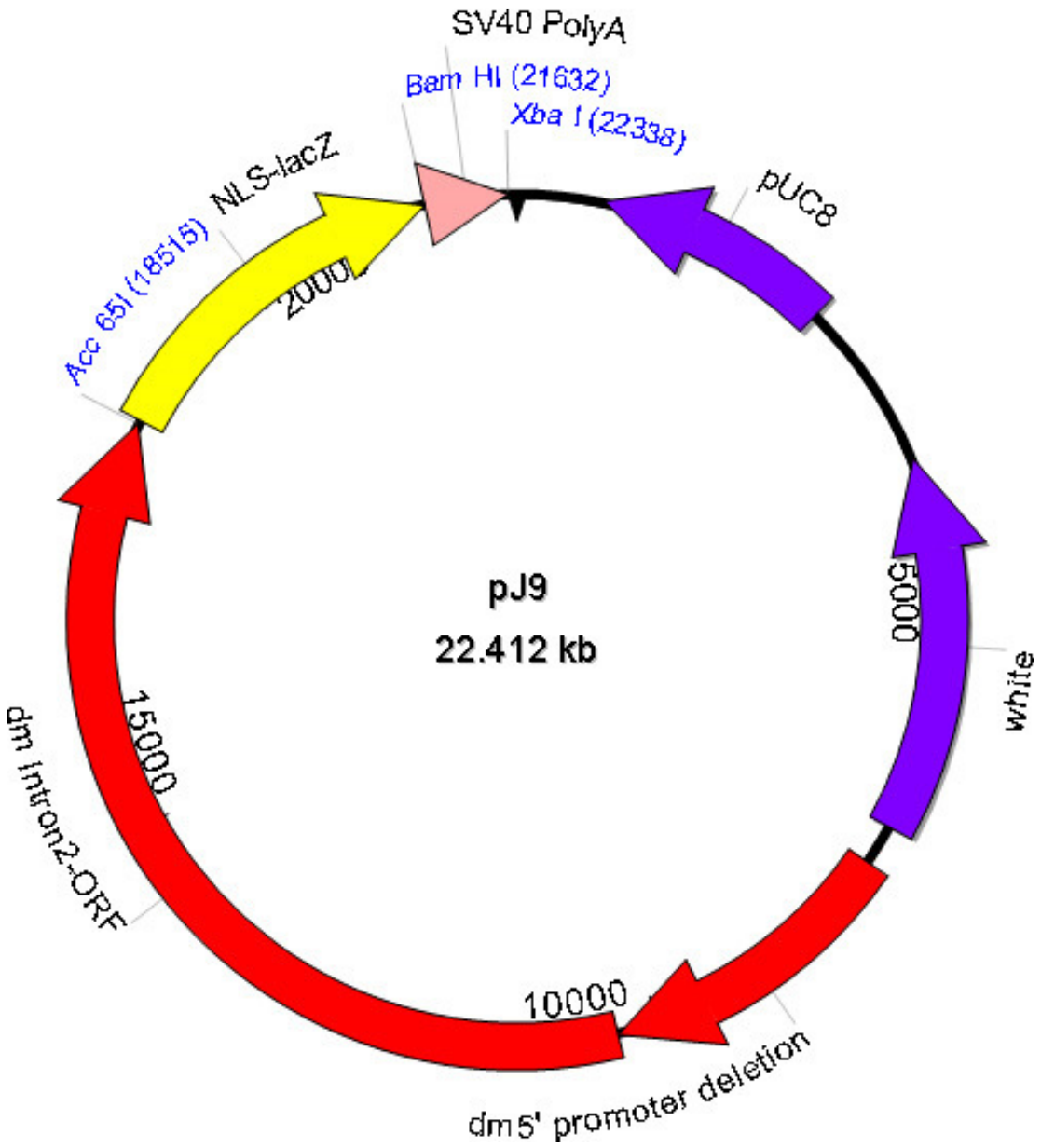



**Figure 5.3.g. J9 plasmid circular map.** In this construct the *dmyc* 5'-end deletion promoter contained in J5 (see Result, Figure 2.4.3.b) is fused to the corrected intron 2 full-length sequence to drive the expression of the *lacZ* gene. The J10 construct (map not shown) is very similar to J9 except that the orientation of the *dmyc* 5' fragment is reverse (3' → 5').

## 5.4 Raw Data

In this study, I have dissected the *dmyc* regulatory regions and have studied the activity of the *lacZ* reporter under control of these regions in a variety of established transgenic lines. The main expression obtained for *lacZ* under control of each transgene is presented in Results sections and interpreted in the Discussion. In this section I shall present raw data for larval brain and imaginal tissues from further transgenic lines that carry sequences from the far upstream, proximal, or intragenic regions. These regions were shown to contain the *dmyc cis*-regulatory elements.

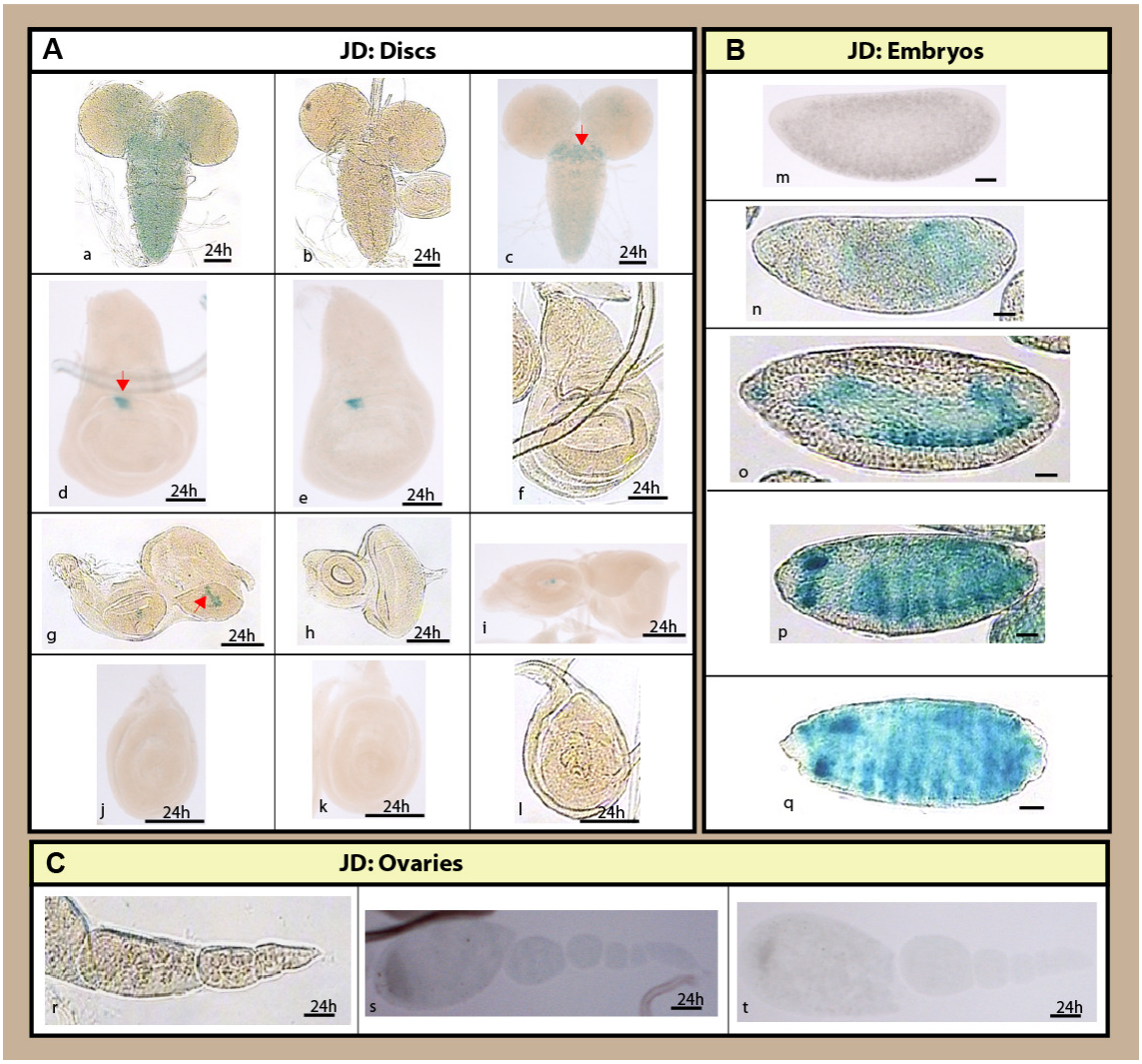



**Figure 5.4.a. Reporter activity under control of far upstream regulatory region (raw data).** (**A**) The JD transgene in JD-42.2 (**a**, **d**, **g**, **j**), JD-61.4 (**b**, **e**, **h**, **k**), and in the JD-68E-21.1 (**c**, **f**, **i**, **l**) was negative in the brain (**a**-**c**) and in imaginal discs (**d**-**f**, wing; **g**-**i**, eye; **j**-**l**, leg discs). (**B**) JD fragment (JD-42.2) was not active in early embryos (**m**, **n**), however, the fragment was active in late embryos (**o**–**q**, stage 13–16). Red arrow in **c** and **d** indicates unspecific background stain deposites, and **g** unspecific staining of macrophages. (**C**) The staining in ovaries for JD-42.2 (**r**), JD-51.1 (**s**), and JD-58.2 (**t**) was negative. Staining time for brains and discs is indicated above the scale bar. Scale bar in (**a**–**t**) indicates 50 µm.

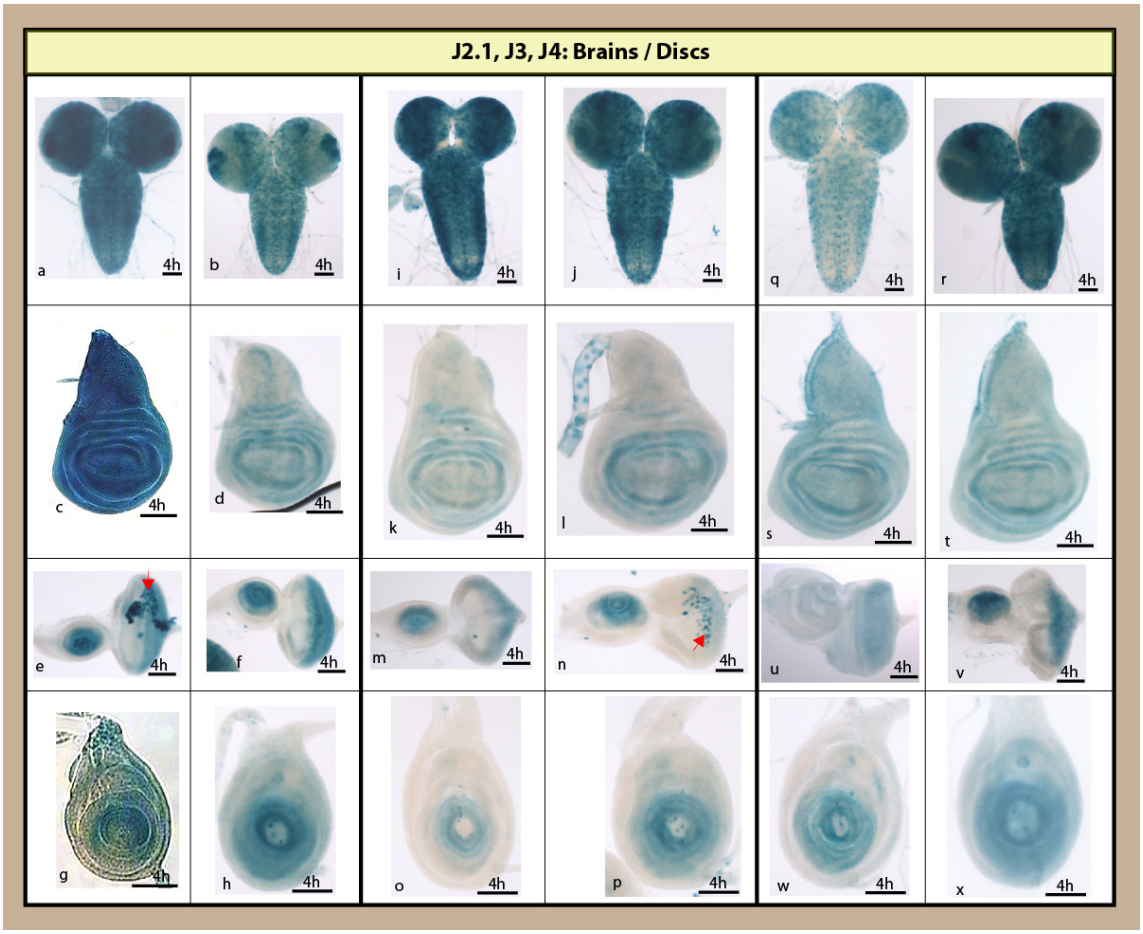



**Figure 5.4.b. *LacZ* activity under control of J2.1-J4 deletions in larval tissues (raw data).** Brains and discs taken from J2.1 (**a**-**h**), J3 (**i**-**p**), and J4 (**q**-**x**) are shown. J2.1 transgene in J2.1-12.1 (**a**, **c**, **e**, **g**), J2.1-86F-1 (**b**, **d**, **f**, **h**), J3 deletion in J3-46.1 (**i**, **l**, **m**), J3-48 (**j**, **n**, **p**), J3-36 (**k**, **o**), J4 fragment in J4-86F-22.1 (**q**, **s**, **w**), J4-86F-17.1 (**u**), and J4-101.6 (**r**, **t**, **v**, **x**) were able to express the reporter in a *dmyc* manner in the brain and discs. Unspecific staining of macrophages in **e** and **n** is indicated by red arrow. Staining time for brains and discs is indicated above the scale bar. Scale bar in (**a**–**x**) indicates 50 µm.

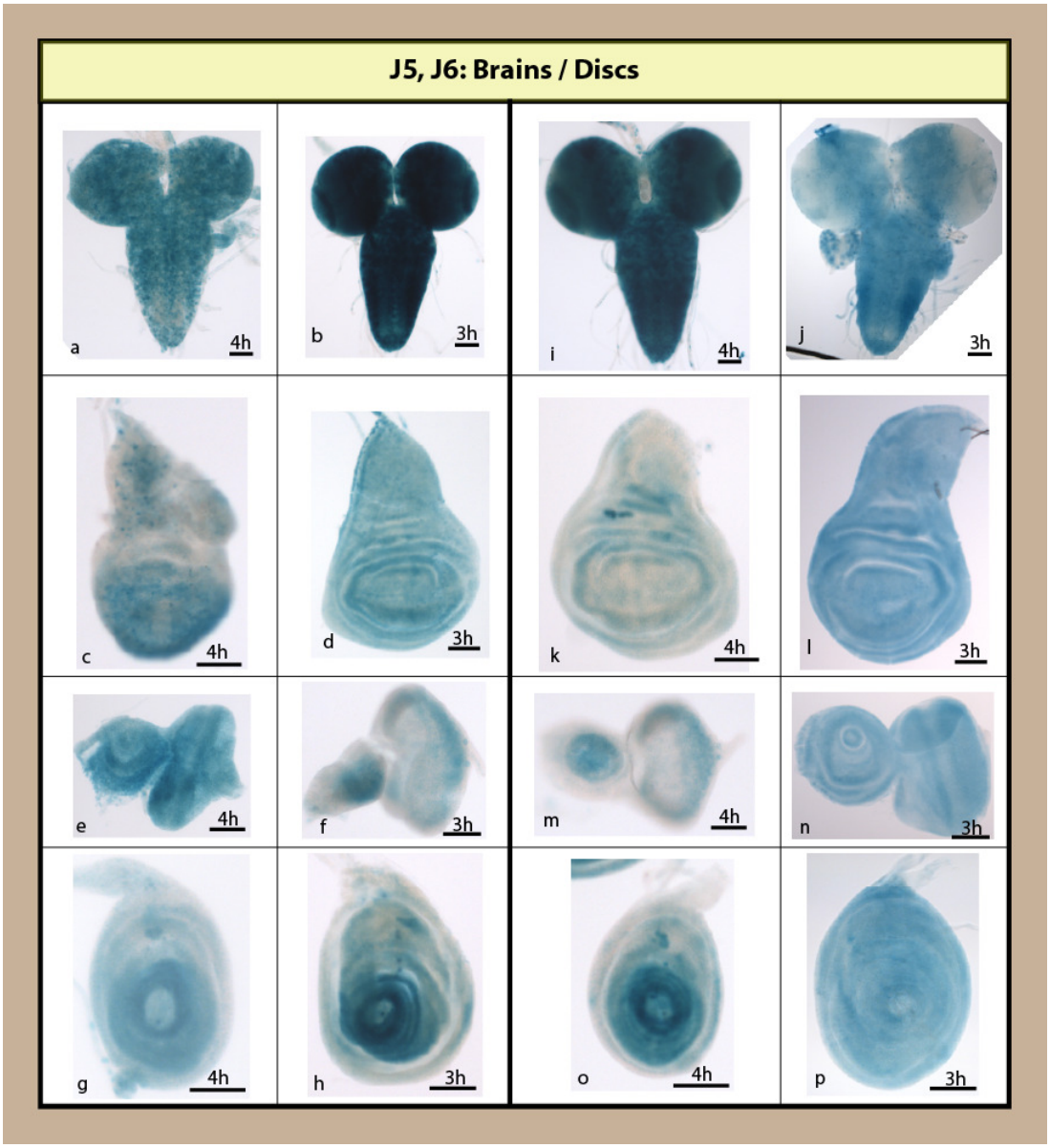



**Figure 5.4.c. *LacZ* activity under control of J5 and J6 transgenes (raw data).** Brains and discs taken from J5 (**a**-**h**) and J6 (**i**-**p**) are shown. J5 truncation in J5-22 (**a**, **c**, **e**, **g**), J5-50.1 (**b**, **f**, **h**), J5-92 (**d**), J6 deletion in J6-34.2 (**i**, **k**, **m**, **o**), and J6-112 (**j**, **l**, **n**, **p**) were able to express the reporter in a *dmyc* manner in the brain and discs. Staining time for brains and discs is indicated above the scale bar. Scale bar in (**a**–**p**) indicates 50 µm.

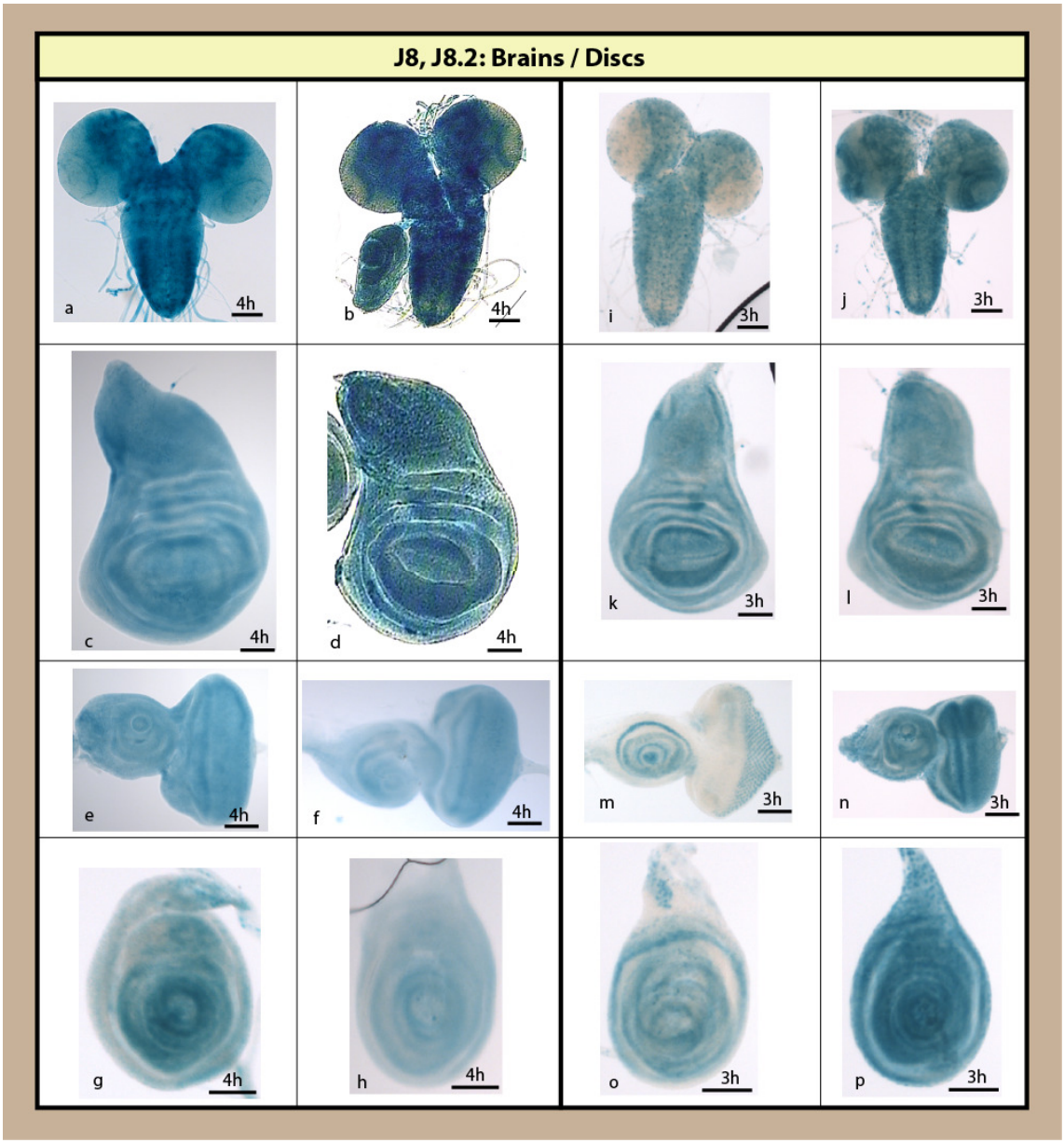



**Figure 5.4.d. *lacZ* expression under control of J8 and J8.2 intronic transgenes (raw data).** The J8 (**a**-**h**) and J8.2 (**i**-**p**) brains and discs are shown. Both transgenes showed *dmyc*-like expression in these tissues: J8- 86F-17.1 (**a**, **g**), J8-32.4 (**c**, **e**), J8-78.1 (**b**, **d**, **f**, **h**), J8.2-17.1 (**i**, **k**, **m**, **o**), and J8.2-59.1 (**j**, **l**, **n**, **p**). Staining time for brains and discs is indicated. Scale bar in (**a**–**p**) indicates 50 µm.

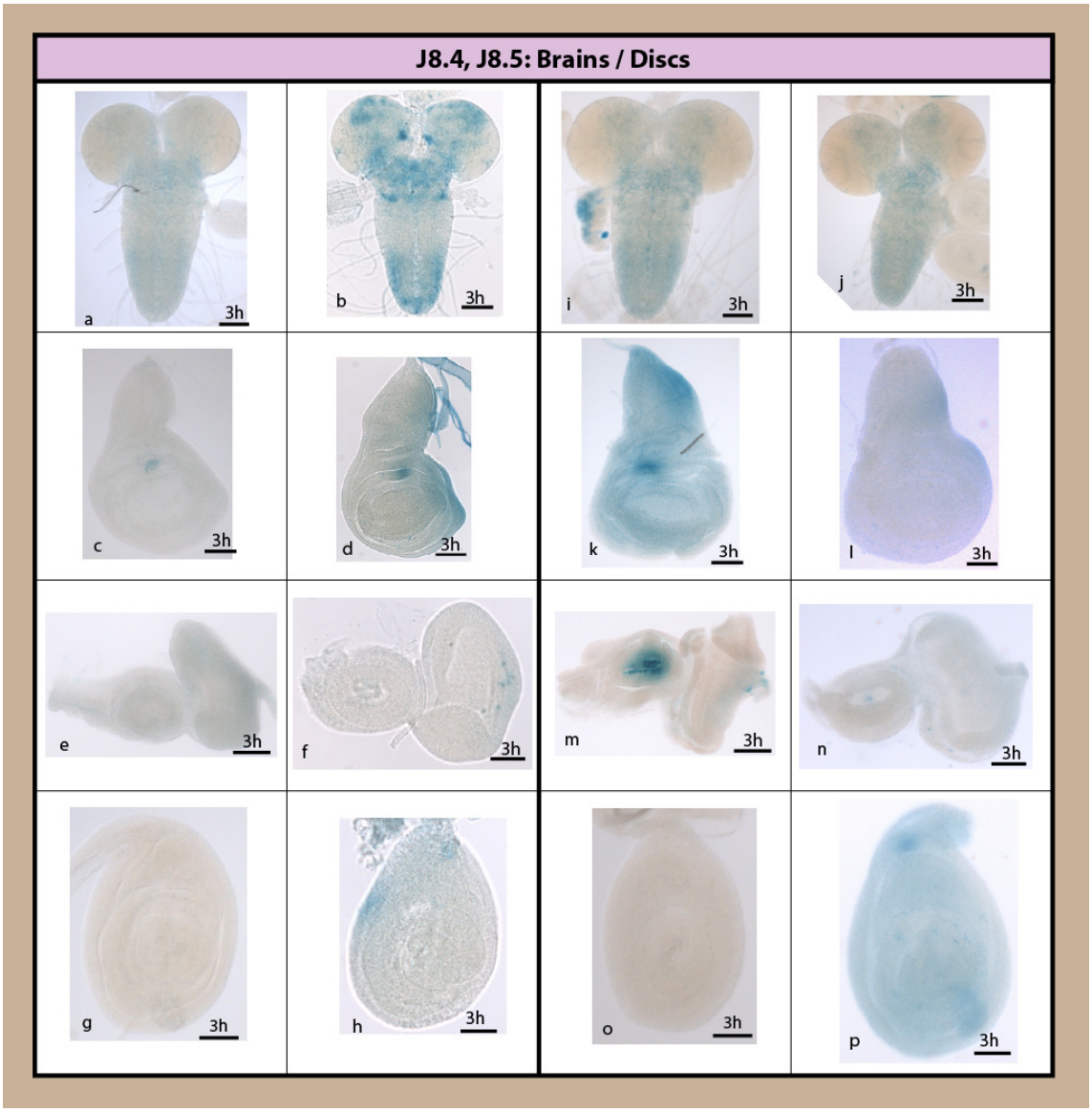



**Figure 5.4.e. *lacZ* expression under control of the intronic deletions J8.4 and J8.5 (raw data).** The J8.4 (**a**-**h**) and the J8.5 (**i**-**p**) brains and discs are shown. J8.4-25710-5 (**a**, **c**, **e**, **g**), J8.4-8622 (**b**, **d**, **f**, **h**), J8.5-25710-5 (**i**, **k**, **m**, **o**), and J8.5-8622 (**j**, **l**, **n**, **p**) were unable to express the reporter beyond the background level in brain and discs (**a**, **b**, **i**, **j** brain; **c**, **d**, **k**, **l** wing discs; **e**, **f**, **m**, **n** eye discs; **g**, **h**, **o**, **p** leg discs). Staining time for brains and discs is indicated above the scale bar. Scale bar in (**a**–**p**) indicates 50 µm.

# 6. References

**Adhikary, S., and Eilers, M. (2005). Transcriptional regulation and transformation by Myc proteins. Nat Rev Mol Cell Biol *6*, 635-645.**

**Aguda, B.D., Kim, Y., Piper-Hunter, M.G., Friedman, A., and Marsh, C.B. (2008). MicroRNA regulation of a cancer network: consequences of the feedback loops involving miR-17-92, E2F, and Myc. Proc Natl Acad Sci U S A *105*, 19678-19683.**

**Amati, B. (2001). Integrating Myc and TGF-beta signalling in cell-cycle control. Nat Cell Biol *3*, E112-113.**

**Amati, B., and Land, H. (1994). Myc-Max-Mad: a transcription factor network controlling cell cycle progression, differentiation and death. Curr Opin Genet Dev *4*, 102-108.**

**Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F.*, et al.* (1981). Sequence and organization of the human mitochondrial genome. Nature *290*, 457-465.**

**Arabi, A., Wu, S., Ridderstrale, K., Bierhoff, H., Shiue, C., Fatyol, K., Fahlen, S., Hydbring, P., Soderberg, O., Grummt, I.*, et al.* (2005). c-Myc associates with ribosomal DNA and activates RNA polymerase I transcription. Nat Cell Biol *7*, 303-310.**

**Aster, J.C. (2005). Deregulated NOTCH signaling in acute T-cell lymphoblastic leukemia/lymphoma: new insights, questions, and opportunities. Int J Hematol *82*, 295-301.**

**Baldwin, R.L., Tran, H., and Karlan, B.Y. (2003). Loss of c-myc repression coincides with ovarian cancer resistance to transforming growth factor beta growth arrest independent of transforming growth factor beta/Smad signaling. Cancer Res *63*, 1413-1419.**

**Bareket-Samish, A., Cohen, I., and Haran, T.E. (2000). Signals for TBP/TATA box recognition. J Mol Biol *299*, 965-977.**

**Bate, M. (1990). The embryonic development of larval muscles in Drosophila. Development *110*, 791-804.**

**Battey, J., Moulding, C., Taub, R., Murphy, W., Stewart, T., Potter, H., Lenoir, G., and Leder, P. (1983). The human c-myc oncogene: structural consequences of translocation into the IgH locus in Burkitt lymphoma. Cell *34*, 779-787.**

**Benassayag, C., Montero, L., Colombie, N., Gallant, P., Cribbs, D., and Morello, D. (2005). Human c-Myc isoforms differentially regulate cell growth and apoptosis in Drosophila melanogaster. Mol Cell Biol *25*, 9897-9909.**

**Bentley, D.L., and Groudine, M. (1986a). A block to elongation is largely responsible for decreased transcription of c-myc in differentiated HL60 cells. Nature *321*, 702-706.**

**Bentley, D.L., and Groudine, M. (1986b). Novel promoter upstream of the human c-myc gene and regulation of c-myc expression in B-cell lymphomas. Mol Cell Biol *6*, 3481-3489.**

**Bieda, M., Xu, X., Singer, M.A., Green, R., and Farnham, P.J. (2006). Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome. Genome Res *16*, 595-605.**

**Bischof, J., Maeda, R.K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proc Natl Acad Sci U S A *104*, 3312-3317.**

**Bissonnette, R.P., Echeverri, F., Mahboubi, A., and Green, D.R. (1992). Apoptotic cell death induced by c-myc is inhibited by bcl-2. Nature *359*, 552-554.**

**Black, A.R., Black, J.D., and Azizkhan-Clifford, J. (2001). Sp1 and kruppel-like factor family of transcription factors in cell growth regulation and cancer. J Cell Physiol *188*, 143-160.**

**Blackwell, T.K., Kretzner, L., Blackwood, E.M., Eisenman, R.N., and Weintraub, H. (1990). Sequence-specific DNA binding by the c-Myc protein. Science *250*, 1149-1151.**

**Blackwood, E.M., and Eisenman, R.N. (1991). Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. Science *251*, 1211-1217.**
**Blackwood, E.M., Kretzner, L., and Eisenman, R.N. (1992). Myc and Max function as a nucleoprotein complex. Curr Opin Genet Dev *2*, 227-235.**
**Bourbon, H.M., Gonzy-Treboul, G., Peronnet, F., Alin, M.F., Ardourel, C., Benassayag, C., Cribbs, D., Deutsch, J., Ferrer, P., Haenlin, M.*, et al.* (2002). A P-insertion screen identifying novel X-linked essential genes in Drosophila. Mech Dev *110*, 71-83.**
**Breathnach, R., and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. Annu Rev Biochem *50*, 349-383.**
**Brewer, G. (1991). An A + U-rich element RNA-binding factor regulates c-myc mRNA stability in vitro. Mol Cell Biol *11*, 2460-2466.**
**Brody, T., Rasband, W., Baler, K., Kuzin, A., Kundu, M., and Odenwald, W.F. (2007). cis-Decoder discovers constellations of conserved DNA sequences shared among tissue-specific enhancers. Genome Biol *8*, R75.**
**Brough, D.E., Hofmann, T.J., Ellwood, K.B., Townley, R.A., and Cole, M.D. (1995). An essential domain of the c-myc protein interacts with a nuclear factor that is also required for E1A-mediated transformation. Mol Cell Biol *15*, 1536-1544.**
**Bueno, M.J., and Malumbres, M. (2011). MicroRNAs and the cell cycle. Biochim Biophys Acta *1812*, 592-601.**
**Bueno, M.J., Perez de Castro, I., and Malumbres, M. (2008). Control of cell proliferation pathways by microRNAs. Cell Cycle *7*, 3143-3148.**
**Burke, T.W., Willy, P.J., Kutach, A.K., Butler, J.E., and Kadonaga, J.T. (1998). The DPE, a conserved downstream core promoter element that is functionally analogous to the TATA box. Cold Spring Harb Symp Quant Biol *63*, 75-82.**
**Cannell, I.G., and Bushell, M. (2010). Regulation of Myc by miR-34c: A mechanism to prevent genomic instability? Cell Cycle *9*, 2726-2730.**
**Carroll, S.B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell *134*, 25-36.**
**Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A.J.*, et al.* (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell *116*, 499-509.**
**Chen, Y., Blackwell, T.W., Chen, J., Gao, J., Lee, A.W., and States, D.J. (2007). Integration of genome and chromatin structure with gene expression profiles to predict c-MYC recognition site binding and function. PLoS Comput Biol *3*, e63.**
**Chin, L., Schreiber-Agus, N., Pellicer, I., Chen, K., Lee, H.W., Dudast, M., Cordon-Cardo, C., and DePinho, R.A. (1995). Contrasting roles for Myc and Mad proteins in cellular growth and differentiation. Proc Natl Acad Sci U S A *92*, 8488-8492.**
**Chung, H.J., and Levens, D. (2005). c-myc expression: keep the noise down! Mol Cells *20*, 157-166.**
**Cleveland, J.L., Huleihel, M., Bressler, P., Siebenlist, U., Akiyama, L., Eisenman, R.N., and Rapp, U.R. (1988). Negative regulation of c-myc transcription involves myc family proteins. Oncogene Res *3*, 357-375.**
**Cohen, B., Wimmer, E.A., and Cohen, S.M. (1991). Early development of leg and wing primordia in the Drosophila embryo. Mech Dev *33*, 229-240.**
**Cole, M.D., and Henriksson, M. (2006). 25 years of the c-Myc oncogene. Semin Cancer Biol *16*, 241.**
**Cotterman, R., Jin, V.X., Krig, S.R., Lemen, J.M., Wey, A., Farnham, P.J., and Knoepfler, P.S. (2008). N-Myc regulates a widespread euchromatic program in the human genome partially independent of its role as a classical transcription factor. Cancer Res *68*, 9654-9662.**
**Cranna, N., and Quinn, L. (2009). Impact of steroid hormone signals on Drosophila cell cycle during development. Cell Div *4*, 3.**

**Cranna, N.J., Mitchell, N.C., Hannan, R.D., and Quinn, L.M. (2011). Hfp, the Drosophila homolog of the mammalian c-myc transcriptional-repressor and tumour suppressor FIR, inhibits dmyc transcription and cell growth. Fly (Austin) *5*.**

**Daines, B., Wang, H., Wang, L., Li, Y., Han, Y., Emmert, D., Gelbart, W., Wang, X., Li, W., Gibbs, R.*, et al.* (2011). The Drosophila melanogaster transcriptome by paired-end RNA sequencing. Genome Res *21*, 315-324.**

**Dang, C.V. (1999). c-Myc target genes involved in cell growth, apoptosis, and metabolism. Mol Cell Biol *19*, 1-11.**

**de la Cova, C., Abril, M., Bellosta, P., Gallant, P., and Johnston, L.A. (2004). Drosophila myc regulates organ size by inducing cell competition. Cell *117*, 107-116.**

**de la Cova, C., and Johnston, L.A. (2006). Myc in model organisms: a view from the flyroom. Semin Cancer Biol *16*, 303-312.**

**Demontis, F., and Perrimon, N. (2009). Integration of Insulin receptor/Foxo signaling and dMyc activity during muscle growth regulates body size in Drosophila. Development *136*, 983-993.**

**Dobens, L.L., and Raftery, L.A. (2000). Integration of epithelial patterning and morphogenesis in Drosophila ovarian follicle cells. Dev Dyn *218*, 80-93.**

**Dogan, R.I., Getoor, L., Wilbur, W.J., and Mount, S.M. (2007). Features generated for computational splice-site prediction correspond to functional elements. BMC Bioinformatics *8*, 410.**

**Dominguez-Sola, D., Ying, C.Y., Grandori, C., Ruggiero, L., Chen, B., Li, M., Galloway, D.A., Gu, W., Gautier, J., and Dalla-Favera, R. (2007). Non-transcriptional control of DNA replication by c-Myc. Nature *448*, 445-451.**

**Down, T.A., Bergman, C.M., Su, J., and Hubbard, T.J. (2007). Large-scale discovery of promoter motifs in Drosophila melanogaster. PLoS Comput Biol *3*, e7.**

**Drummond, D.R., McCrae, M.A., and Colman, A. (1985). Stability and movement of mRNAs and their encoded proteins in Xenopus oocytes. J Cell Biol *100*, 1148-1156.**

**Duman-Scheel, M., Johnston, L.A., and Du, W. (2004). Repression of dMyc expression by Wingless promotes Rbf-induced G1 arrest in the presumptive Drosophila wing margin. Proc Natl Acad Sci U S A *101*, 3857-3862.**

**Dvir, A., Conaway, J.W., and Conaway, R.C. (2001). Mechanism of transcription initiation and promoter escape by RNA polymerase II. Curr Opin Genet Dev *11*, 209-214.**

**Eilers, M., and Eisenman, R.N. (2008). Myc's broad reach. Genes Dev *22*, 2755-2766.**

**Eisenman, R.N. (2001). Deconstructing myc. Genes Dev *15*, 2023-2030.**

**Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D.S. (2003). MicroRNA targets in Drosophila. Genome Biol *5*, R1.**

**Evan, G.I., and Littlewood, T.D. (1993). The role of c-myc in cell growth. Curr Opin Genet Dev *3*, 44-49.**

**Fanidi, A., Harrington, E.A., and Evan, G.I. (1992). Cooperative interaction between c-myc and bcl-2 proto-oncogenes. Nature *359*, 554-556.**

**Gallant, P., Shiio, Y., Cheng, P.F., Parkhurst, S.M., and Eisenman, R.N. (1996). Myc and Max homologs in Drosophila. Science *274*, 1523-1527.**

**Gao, K., Masuda, A., Matsuura, T., and Ohno, K. (2008). Human branch point consensus sequence is yUnAy. Nucleic Acids Res *36*, 2257-2267.**

**Gasch, A., Hinz, U., and Renkawitz-Pohl, R. (1989). Intron and upstream sequences regulate expression of the Drosophila beta 3-tubulin gene in the visceral and somatic musculature, respectively. Proc Natl Acad Sci U S A *86*, 3215-3218.**

**Gombert, W.M., Farris, S.D., Rubio, E.D., Morey-Rosler, K.M., Schubach, W.H., and Krumm, A. (2003). The c-myc insulator element and matrix attachment regions define the c-myc chromosomal domain. Mol Cell Biol *23*, 9338-9348.**

**Goodliffe, J.M., Wieschaus, E., and Cole, M.D. (2005). Polycomb mediates Myc autorepression and its transcriptional control of many loci in Drosophila. Genes Dev *19*, 2941-2946.**

**Grandori, C., Cowley, S.M., James, L.P., and Eisenman, R.N. (2000). The Myc/Max/Mad network and the transcriptional control of cell behavior. Annu Rev Cell Dev Biol *16*, 653-699.**
**Green, M.R. (1986). Pre-mRNA splicing. Annu Rev Genet *20*, 671-708.**
**Green, M.R. (1991). Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. Annu Rev Cell Biol *7*, 559-599.**
**Grewal, S.S., Li, L., Orian, A., Eisenman, R.N., and Edgar, B.A. (2005). Myc-dependent regulation of ribosomal RNA synthesis during Drosophila development. Nat Cell Biol *7*, 295-302.**
**Gross, P., and Oelgeschlager, T. (2006). Core promoter-selective RNA polymerase II transcription. Biochem Soc Symp, 225-236.**
**Guo, Y., Niu, C., Breslin, P., Tang, M., Zhang, S., Wei, W., Kini, A.R., Paner, G.P., Alkan, S., Morris, S.W*., et al.* (2009). c-Myc-mediated control of cell fate in megakaryocyte-erythrocyte progenitors. Blood *114*, 2097-2106.**
**Hallikas, O., Palin, K., Sinjushina, N., Rautiainen, R., Partanen, J., Ukkonen, E., and Taipale, J. (2006). Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. Cell *124*, 47-59.**
**Harris, R.E., Pargett, M., Sutcliffe, C., Umulis, D., and Ashe, H.L. (2011). Brat promotes stem cell differentiation via control of a bistable switch that restricts BMP signaling. Dev Cell *20*, 72-83.**
**He, T.C., Sparks, A.B., Rago, C., Hermeking, H., Zawel, L., da Costa, L.T., Morin, P.J., Vogelstein, B., and Kinzler, K.W. (1998). Identification of c-MYC as a target of the APC pathway. Science *281*, 1509-1512.**
**He, W., Dai, C., Li, Y., Zeng, G., Monga, S.P., and Liu, Y. (2009). Wnt/beta-catenin signaling promotes renal interstitial fibrosis. J Am Soc Nephrol *20*, 765-776.**
**Henriksson, M., and Luscher, B. (1996). Proteins of the Myc network: essential regulators of cell growth and differentiation. Adv Cancer Res *68*, 109-182.**
**Herkert, B., and Eilers, M. (2011). Transcriptional repression: the dark side of myc. Genes Cancer *1*, 580-586.**
**Hermeking, H., and Eick, D. (1994). Mediation of c-Myc-induced apoptosis by p53. Science *265*, 2091-2093.**
**Herranz, H., Hong, X., Perez, L., Ferreira, A., Olivieri, D., Cohen, S.M., and Milan, M. (2010). The miRNA machinery targets Mei-P26 and regulates Myc protein levels in the Drosophila wing. EMBO J *29*, 1688-1698.**
**Huang, Y., and Carmichael, G.G. (1996). A suboptimal 5' splice site is a cis-acting determinant of nuclear export of polyomavirus late mRNAs. Mol Cell Biol *16*, 6046-6054.**
**Ishihara, K., Oshimura, M., and Nakao, M. (2006). CTCF-dependent chromatin insulator is linked to epigenetic remodeling. Mol Cell *23*, 733-742.**
**Jaglarz, M.K., Krzeminski, W., and Bilinski, S.M. (2008). Structure of the ovaries and follicular epithelium morphogenesis in Drosophila and its kin. Dev Genes Evol *218*, 399-411.**
**Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B., and Smale, S.T. (1994). DNA sequence requirements for transcriptional initiator activity in mammalian cells. Mol Cell Biol *14*, 116-127.**
**Jeibmann, A., and Paulus, W. (2009). Drosophila melanogaster as a model organism of brain diseases. Int J Mol Sci *10*, 407-440.**
**Johnston, L.A., and Edgar, B.A. (1998). Wingless and Notch regulate cell-cycle arrest in the developing Drosophila wing. Nature *394*, 82-84.**
**Johnston, L.A., Prober, D.A., Edgar, B.A., Eisenman, R.N., and Gallant, P. (1999). Drosophila myc regulates cellular growth during development. Cell *98*, 779-790.**
**Jones, R.M., Branda, J., Johnston, K.A., Polymenis, M., Gadd, M., Rustgi, A., Callanan, L., and Schmidt, E.V. (1996). An essential E box in the promoter of the gene encoding the mRNA cap-binding protein (eukaryotic initiation factor 4E) is a target for activation by c-myc. Mol Cell Biol *16*, 4754-4764.**
**Juven-Gershon, T., Hsu, J.Y., and Kadonaga, J.T. (2006). Perspectives on the RNA polymerase II core promoter. Biochem Soc Trans *34*, 1047-1050.**

**Kelloff, G., and Vogt, P.K. (1966). Localization of avian tumor virus group-specific antigen in cell and virus. Virology *29*, 377-384.**
**Kerkhoff, E., Houben, R., Loffler, S., Troppmair, J., Lee, J.E., and Rapp, U.R. (1998). Regulation of c-myc expression by Ras/Raf signalling. Oncogene *16*, 211-216.**
**Kharazmi, J., Moshfegh, C., and Brody, T. (2011). Identification of cis-Regulatory Elements in the dmyc Gene of Drosophila Melanogaster. Gene Regulation and Systems Biology *6*, 15.**
**Kiernan, J.A. (2007). Indigogenic substrates for detection and localization of enzymes. Biotech Histochem *82*, 73-103.**
**King, M.W., Roberts, J.M., and Eisenman, R.N. (1986). Expression of the c-myc proto-oncogene during development of Xenopus laevis. Mol Cell Biol *6*, 4499-4508.**
**Kipp, J.L., and Mayo, K.E. (2009). Use of reporter genes to study the activity of promoters in ovarian granulosa cells. Methods Mol Biol *590*, 177-193.**
**Knudsen, K.J., Nelander Holm, G.M., Krabbe, J.S., Listov-Saabye, N., Kiehr, B., Dufva, M., Svendsen, J.E., and Oleksiewicz, M.B. (2009). Driving gradual endogenous c-myc overexpression by flow-sorting: intracellular signaling and tumor cell phenotype correlate with oncogene expression. Arch Toxicol.**
**Knudsen, S. (1999). Promoter2.0: for the recognition of PolII promoter sequences. Bioinformatics *15*, 356-361.**
**Kollmar, R., and Farnham, P.J. (1993). Site-specific initiation of transcription by RNA polymerase II. Proc Soc Exp Biol Med *203*, 127-139.**
**Krug, M.S., and Berger, S.L. (1987). First-strand cDNA synthesis primed with oligo(dT). Methods Enzymol *152*, 316-325.**
**Kulkarni, M.M., and Arnosti, D.N. (2005). cis-regulatory logic of short-range transcriptional repression in Drosophila melanogaster. Mol Cell Biol *25*, 3411-3420.**
**Kutach, A.K., and Kadonaga, J.T. (2000). The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. Mol Cell Biol *20*, 4754-4764.**
**Lang, J.C., Whitelaw, B., Talbot, S., and Wilkie, N.M. (1988). Transcriptional regulation of the human c-myc gene. Br J Cancer Suppl *9*, 62-66.**
**Leder, P., Battey, J., Lenoir, G., Moulding, C., Murphy, W., Potter, H., Stewart, T., and Taub, R. (1983). Translocations among antibody genes in human cancer. Science *222*, 765-771.**
**Lee, T.I., and Young, R.A. (1998). Regulation of gene expression by TBP-associated proteins. Genes Dev *12*, 1398-1408.**
**Lemm, I., and Ross, J. (2002). Regulation of c-myc mRNA decay by translational pausing in a coding region instability determinant. Mol Cell Biol *22*, 3959-3969.**
**Levens, D.L. (2003). Reconstructing MYC. Genes Dev *17*, 1071-1077.**
**Lifton, R.P., Goldberg, M.L., Karp, R.W., and Hogness, D.S. (1978). The organization of the histone genes in Drosophila melanogaster: functional and evolutionary implications. Cold Spring Harb Symp Quant Biol *42 Pt 2*, 1047-1051.**
**Liu, H., Zhou, M., Luo, X., Zhang, L., Niu, Z., Peng, C., Ma, J., Peng, S., Zhou, H., Xiang, B.*, et al.* (2008). Transcriptional regulation of BRD7 expression by Sp1 and c-Myc. BMC Mol Biol *9*, 111.**
**Liu, J., Kouzine, F., Nie, Z., Chung, H.J., Elisha-Feil, Z., Weber, A., Zhao, K., and Levens, D. (2006). The FUSE/FBP/FIR/TFIIH system is a molecular machine programming a pulse of c-myc expression. EMBO J *25*, 2119-2130.**
**Liu, J., and Levens, D. (2006). Making myc. Curr Top Microbiol Immunol *302*, 1-32.**
**Lo, K., and Smale, S.T. (1996). Generality of a functional initiator consensus sequence. Gene *182*, 13-22.**
**Luscher, B., and Eisenman, R.N. (1990). New light on Myc and Myb. Part I. Myc. Genes Dev *4*, 2025-2035.**
**Lutz, C.S. (2008). Alternative polyadenylation: a twist on mRNA 3' end formation. ACS Chem Biol *3*, 609-617.**
**Maillard, I., Fang, T., and Pear, W.S. (2005). Regulation of lymphoid development, differentiation, and function by the Notch pathway. Annu Rev Immunol *23*, 945-974.**

**Marcu, K.B. (1987). Regulation of expression of the c-myc proto-oncogene. Bioessays *6*, 28-32.**
**Marcu, K.B., Bossone, S.A., and Patel, A.J. (1992). myc function and regulation. Annu Rev Biochem *61*, 809-860.**
**Mautner, J., Behrends, U., Hortnagel, K., Brielmeier, M., Hammerschmidt, W., Strobl, L., Bornkamm, G.W., and Polack, A. (1996). c-myc expression is activated by the immunoglobulin kappa-enhancers from a distance of at least 30 kb but not by elements located within 50 kb of the unaltered c-myc locus in vivo. Oncogene *12*, 1299-1307.**
**Mautner, J., Joos, S., Werner, T., Eick, D., Bornkamm, G.W., and Polack, A. (1995). Identification of two enhancer elements downstream of the human c-myc gene. Nucleic Acids Res *23*, 72-80.**
**McQuilton, P., St Pierre, S.E., and Thurmond, J. (2012). FlyBase 101--the basics of navigating FlyBase. Nucleic Acids Res *40*, D706-714.**
**Mitchell, N.C., Johanson, T.M., Cranna, N.J., Er, A.L., Richardson, H.E., Hannan, R.D., and Quinn, L.M. (2010). Hfp inhibits Drosophila myc transcription and cell growth in a TFIIH/Hay-dependent manner. Development.**
**Mount, S.M., Burks, C., Hertz, G., Stormo, G.D., White, O., and Fields, C. (1992). Splicing signals in Drosophila: intron size, information content, and consensus sequences. Nucleic Acids Res *20*, 4255-4262.**
**Murphy, L.C., Huzel, N., and Davie, J.R. (1996). Novel DNase I hypersensitive sites in the 3'-flanking region of the human c-myc gene. DNA Cell Biol *15*, 543-548.**
**Nau, M.M., Brooks, B.J., Battey, J., Sausville, E., Gazdar, A.F., Kirsch, I.R., McBride, O.W., Bertness, V., Hollis, G.F., and Minna, J.D. (1985). L-myc, a new myc-related gene amplified and expressed in human small cell lung cancer. Nature *318*, 69-73.**
**Neto-Silva, R.M., de Beco, S., and Johnston, L.A. (2010). Evidence for a growth-stabilizing regulatory feedback mechanism between Myc and Yorkie, the Drosophila homolog of Yap. Dev Cell *19*, 507-520.**
**Orian, A., Delrow, J.J., Rosales Nieves, A.E., Abed, M., Metzger, D., Paroush, Z., Eisenman, R.N., and Parkhurst, S.M. (2007). A Myc-Groucho complex integrates EGF and Notch signaling to regulate neural development. Proc Natl Acad Sci U S A *104*, 15771-15776.**
**Orian, A., van Steensel, B., Delrow, J., Bussemaker, H.J., Li, L., Sawado, T., Williams, E., Loo, L.W., Cowley, S.M., Yost, C*., et al.* (2003). Genomic binding by the Drosophila Myc, Max, Mad/Mnt transcription factor network. Genes Dev *17*, 1101-1114.**
**Oster, S.K., Ho, C.S., Soucie, E.L., and Penn, L.Z. (2002). The myc oncogene: MarvelouslY Complex. Adv Cancer Res *84*, 81-154.**
**Patel, J.H., and McMahon, S.B. (2006). Targeting of Miz-1 is essential for Myc-mediated apoptosis. J Biol Chem *281*, 3283-3289.**
**Popescu, N.C., and Zimonjic, D.B. (2002). Chromosome-mediated alterations of the MYC gene in human cancer. J Cell Mol Med *6*, 151-159.**
**Preiss, T., and Hentze, M.W. (1998). Dual function of the messenger RNA cap structure in poly(A)-tail-promoted translation in yeast. Nature *392*, 516-520.**
**Prendergast, G.C. (1999). Mechanisms of apoptosis by c-Myc. Oncogene *18*, 2967-2987.**
**Qi, L., Haurwitz, R.E., Shao, W., Doudna, J.A., and Arkin, A.P. (2012). RNA processing enables predictable programming of gene expression. Nat Biotechnol.**
**Ray, D., and Robert-Lezenes, J. (1989). Coexistence of a c-myc mRNA initiated in intron 1 with the normal c-myc mRNA and similar regulation of both transcripts in mammalian cells. Oncogene Res *5*, 73-78.**
**Reneker, J., Lyons, E., Conant, G.C., Pires, J.C., Freeling, M., Shyu, C.R., and Korkin, D. (2012). Long identical multispecies elements in plant and animal genomes. Proc Natl Acad Sci U S A *109*, E1183-1191.**
**Rivera-Pomar, R., and Jackle, H. (1996). From gradients to stripes in Drosophila embryogenesis: filling in the gaps. Trends Genet *12*, 478-483.**

**Roy, S., Ernst, J., Kharchenko, P.V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L., Lin, M.F.*, et al.* (2010). Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science *330*, 1787-1797.**

**Rubin, G.M. (1988). Drosophila melanogaster as an experimental organism. Science *240*, 1453-1459.**

**Rubin, G.M., and Spradling, A.C. (1983). Vectors for P element-mediated gene transfer in Drosophila. Nucleic Acids Res *11*, 6341-6351.**

**Sakamuro, D., and Prendergast, G.C. (1999). New Myc-interacting proteins: a second Myc network emerges. Oncogene *18*, 2942-2954.**

**Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., and Hume, D.A. (2007). Mammalian RNA polymerase II core promoters: insights from genome-wide studies. Nat Rev Genet *8*, 424-436.**

**Schmidt, E.V. (1999). The role of c-myc in cellular growth control. Oncogene *18*, 2988-2996.**

**Schreiber-Agus, N., Alland, L., Muhle, R., Goltz, J., Chen, K., Stevens, L., Stein, D., and DePinho, R.A. (1997). A biochemical and biological analysis of Myc superfamily interactions. Curr Top Microbiol Immunol *224*, 159-168.**

**Sears, R., Leone, G., DeGregori, J., and Nevins, J.R. (1999). Ras enhances Myc protein stability. Mol Cell *3*, 169-179.**

**Siddall, N.A., Lin, J.I., Hime, G.R., and Quinn, L.M. (2009). Myc--what we have learned from flies. Curr Drug Targets *10*, 590-601.**

**Simcox, A.A., Hersperger, E., Shearn, A., Whittle, J.R., and Cohen, S.M. (1991). Establishment of imaginal discs and histoblast nests in Drosophila. Mech Dev *34*, 11-20.**

**Smale, S.T. (2001). Core promoters: active contributors to combinatorial gene regulation. Genes Dev *15*, 2503-2508.**

**Smale, S.T., Jain, A., Kaufmann, J., Emami, K.H., Lo, K., and Garraway, I.P. (1998). The initiator element: a paradigm for core promoter heterogeneity within metazoan protein-coding genes. Cold Spring Harb Symp Quant Biol *63*, 21-31.**

**Smale, S.T., and Kadonaga, J.T. (2003). The RNA polymerase II core promoter. Annu Rev Biochem *72*, 449-479.**

**Sodir, N.M., and Evan, G.I. (2009). Nursing some sense out of Myc. J Biol *8*, 77.**

**Song, H., Hasson, P., Paroush, Z., and Courey, A.J. (2004). Groucho oligomerization is required for repression in vivo. Mol Cell Biol *24*, 4341-4350.**

**Song, Y., Wu, J., Oyesanya, R.A., Lee, Z., Mukherjee, A., and Fang, X. (2009). Sp-1 and c-Myc mediate lysophosphatidic acid-induced expression of vascular endothelial growth factor in ovarian cancer cells via a hypoxia-inducible factor-1-independent mechanism. Clin Cancer Res *15*, 492-501.**

**Spencer, C.A., and Groudine, M. (1991). Control of c-myc regulation in normal and neoplastic cells. Adv Cancer Res *56*, 1-48.**

**Spotts, G.D., Patel, S.V., Xiao, Q., and Hann, S.R. (1997). Identification of downstream-initiated c-Myc proteins which are dominant-negative inhibitors of transactivation by full-length c-Myc proteins. Mol Cell Biol *17*, 1459-1468.**

**Stewart, T.A., Bellve, A.R., and Leder, P. (1984). Transcription and promoter usage of the myc gene in normal somatic and spermatogenic cells. Science *226*, 707-710.**

**Taghavi, P., Verhoeven, E., Jacobs, J.J., Lambooij, J.P., Stortelers, C., Tanger, E., Moolenaar, W.H., and van Lohuizen, M. (2008). In vitro genetic screen identifies a cooperative role for LPA signaling and c-Myc in cell transformation. Oncogene *27*, 6806-6816.**

**Taylor, A.C., Schuster, K., McKenzie, P.P., and Harris, L.C. (2006). Differential cooperation of oncogenes with p53 and Bax to induce apoptosis in rhabdomyosarcoma. Mol Cancer *5*, 53.**

**Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res *22*, 4673-4680.**

**Trumpp, A., Refaeli, Y., Oskarsson, T., Gasser, S., Murphy, M., Martin, G.R., and Bishop, J.M. (2001). c-Myc regulates mammalian body size by controlling cell number but not cell size. Nature *414*, 768-773.**

**Venken, K.J., He, Y., Hoskins, R.A., and Bellen, H.J. (2006). P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster. Science *314*, 1747-1751.**

**Wachtel, C., and Manley, J.L. (2009). Splicing of mRNA precursors: the role of RNAs and proteins in catalysis. Mol Biosyst *5*, 311-316.**

**Watt, R., Nishikura, K., Sorrentino, J., ar-Rushdi, A., Croce, C.M., and Rovera, G. (1983). The structure and nucleotide sequence of the 5' end of the human c-myc oncogene. Proc Natl Acad Sci U S A *80*, 6307-6311.**

**Weinberg, R.A. (1995). The retinoblastoma protein and cell cycle control. Cell *81*, 323-330.**

**Welstead, G.G., Brambrink, T., and Jaenisch, R. (2008). Generating iPS cells from MEFS through forced expression of Sox-2, Oct-4, c-Myc, and Klf4. J Vis Exp.**

**Weng, A.P., Millholland, J.M., Yashiro-Ohtani, Y., Arcangeli, M.L., Lau, A., Wai, C., Del Bianco, C., Rodriguez, C.G., Sai, H., Tobias, J.*, et al.* (2006). c-Myc is an important direct target of Notch1 in T-cell acute lymphoblastic leukemia/lymphoma. Genes Dev *20*, 2096-2109.**

**Wierstra, I., and Alves, J. (2008). The c-myc promoter: still MysterY and challenge. Adv Cancer Res *99*, 113-333.**

**Will, C.L., and Luhrmann, R. (2006). Spliceosome structure and function. Cold Spring Harb Perspect Biol *3*.**

**Yap, C.S., Peterson, A.L., Castellani, G., Sedivy, J.M., and Neretti, N. (2011). Kinetic profiling of the c-Myc transcriptome and bioinformatic analysis of repressed gene promoters. Cell Cycle *10*, 2184-2196.**

**Yavatkar, A.S., Lin, Y., Ross, J., Fann, Y., Brody, T., and Odenwald, W.F. (2008). Rapid detection and curation of conserved DNA via enhanced-BLAT and EvoPrinterHD analysis. BMC Genomics *9*, 106.**

**Yean, D., and Gralla, J. (1997). Transcription reinitiation rate: a special role for the TATA box. Mol Cell Biol *17*, 3809-3816.**

**Zeitlinger, J., and Stark, A. (2010). Developmental gene regulation in the era of genomics. Dev Biol *339*, 230-239.**

**Zeller, K.I., Zhao, X., Lee, C.W., Chiu, K.P., Yao, F., Yustein, J.T., Ooi, H.S., Orlov, Y.L., Shahab, A., Yong, H.C.*, et al.* (2006). Global mapping of c-Myc binding sites and target gene networks in human B cells. Proc Natl Acad Sci U S A *103*, 17834-17839.**

**Zhao, J., Hyman, L., and Moore, C. (1999). Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. Microbiol Mol Biol Rev *63*, 405-445.**

**Zhu, Y.Y., Machleder, E.M., Chenchik, A., Li, R., and Siebert, P.D. (2001). Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. Biotechniques *30*, 892-897.**

**Zimmerman, K.A., Yancopoulos, G.D., Collum, R.G., Smith, R.K., Kohl, N.E., Denis, K.A., Nau, M.M., Witte, O.N., Toran-Allerand, D., Gee, C.E.*, et al.* (1986). Differential expression of myc family genes during murine development. Nature *319*, 780-783.**

## Acknowledgements

This work was supported through different sources. I invaluably appreciate all the contributions which were vital during all the years of my thesis to continue the project. It has been a great pleasure for me to collaborate and interact with the following people, to whom I heartedly send my special thanks:

Prof. Dr. Eric Kubli encouraged me with the continuation of my studies through all the difficult phases of the work. Despite his retirement, he took many hours of his time to read through my work and gave critical and constructive advice in looking for the appropriate corrections, and structuring the work scientifically. I am thankful for the free atmosphere in working with him that enabled me to ask all my questions with no hampering, and he answered them with composure.

Prof. Dr. Hans-Peter Lipp accepted the commitment of being my Thesis Committee chairman, which enabled me to enroll into graduate school program. I am especially grateful, for the unconventional support and confidence he showed to my person.

I am grateful to Prof. Dr. François Karch from the University of Geneva. He made it possible for me to write my Thesis by joining the committee as active Drosophilist. He and members of his laboratory, in particular Dr. Robert Maeda and Dragan Gligorov, gave me many ideas in performing the experiments. Prof. Karch conducted my work at its final phase, and he read and discussed my manuscript critically.

With the commitment of Prof. Dr. Bernard Schmid, my committee board was complete, and I could see the hope for graduation. I am grateful to his contributions in legal matters with the faculty Dean Office.

Dr. Daniel Bopp joined my committee as a Drosophilist and Molecular Biologist from the University of Zurich, Institute for Molecular Life Sciences. I invaluably appreciate his support in reading and discussing my manuscripts critically, and also giving me a good deal of advice with the 5' RACE experiment and genetic crosses.

My son Cameron supported me with the lab work in Zurich-Schlieren. He did a great deal of work with oligonucleotide designing, restriction digestions, and plasmid DNA preparations. With his knowledge in Physics he created an electroelution method with which we could gel extract the large DNA fragment of 27-kb harboring the *dmyc* gene. I am greatly thankful for his support.

Dr. Kevin Cook supported me with all the genetic crosses and orders of Fly Stocks at the Bloomington Stock Center. His continuous help deserves my greatest thankfulness.

Dr. Thomas Brody made it possible to search the *dmyc* locus with their *Evo*-Printer/ *cis*-Decoder Soft Ware for the existence of conserved sequence blocks in noncoding regions. Jean Marie Smith continuously supported me with the use of different

modules of LASERGENE DNASTAR 9.1 in performing Bioinformatics Analyses. I appreciate all the support that I received with bioinformatics work.

Last but not least, I am grateful to Mario Jenni, president of Bio-Technopark Zurich, who enabled me setting up a lab place in the Center to conduct my experiments.

**Jasmine Kharazmi** **Curriculum Vitae**

## EDUCATION

2008 **PhD Candidate** Molecular Biology, University of Zurich

**Dissertation Committee Members,** Prof. F. Karch, University of Geneva; Prof. H-P. Lipp, Prof. B. Schmid, Dr. D. Bopp, University of Zurich

**Dissertation Title**: Identification of Regulatory Regions in the *Drosophila dmyc* Gene: Bioinformatics Analyses Combined with Reporter Activity Studies

2001 **M.S. Molecular Biology**, University of Zurich, Prof. Walter Schaffner

1983 **B.S. Biology**, American Organized University of Shiraz, Iran

## FURTHER EDUCATION

2008 **Advanced *Drosophila* Genetics Course**, Shanghai (China), Prof. Michael Ashburner, Cambridge, Prof. Scott Hawling Stowers Institute for Medical Research, Kansas City, Prof. Casey Bergman, University of Manchester UK

2004 **Handling Biohazard**, Swiss Federal Environmental Department (BUWAL)

1992 **Working with and Handling Radioactive Substances Related to Biomedical Research**, Paul Scherrer Institute/Swiss Federal Institute of Technology Zurich, Switzerland

1988 – 1991 **Diploma** Multivalent Diagnostic Laboratory, University Hospital Zurich and Swiss Red Cross Organization

## RESEARCH / RESEARECH-RELATED EXPERIENCE

2005 – 2012 **Founder and Researcher** „Molecular Biology Lab", affiliated to the University of Zurich, Non-Profit Research Lab at Bio-Technopark Zurich-Schlieren,
Aim: Molecular Biological Basic Research with the Model System Fruit Fly (Pursuing Accomplishment of PhD at the University of Zurich)

2007 – 2011 **Medical Technical Assistant,** Hospital Diagnostic Laboratories Zurich
Aim: Funding my Research and PhD

2003 **Research assistant**, University of Zurich, Laboratory of Prof. Hans-Peter Lipp, Collaboration on the Project „Learning & Memory"

2002 **Research Assistant**, University of Bern, Laboratory of Prof. Hans-Peter Marti, Project „Targeted Drug and Polynucleotide Delivery for the Treatment of Patients with Improved Renal Insufficiency"

1999 **Undergraduate Summer Practical Work**, University of Zurich, Laboratory of Prof. Walter Schaffner, Genotyping of Transgenic Mice

1991 – 1998 **Medical Technical Assistant,** Different Hospital Diagnostic Laboratories, Zurich

## TEACHING EXPERIENCE

1980 – 1983 **Teacher, Biology**, Teacher's Training Center, Shiraz, Iran, and **Instructor**, Freshman Biology Students, University of Shiraz, Iran

## PROFESSIONAL SOCIETY MEMBERSHIP

2012 – present **American Cancer Society (ACS)**
2005 – present **Genetics Society of America (GSA)**

## PUBLICATIONS

2013 **Kharazmi, J.** and Moshfegh, C. Investigation of *dmyc* Promoter and Regulatory Regions (Accepted)

2011 **Kharazmi, J.**, Moshfegh, C., Brody, T. Identification of *cis*-regulatory elements in the *dmyc* gene of *Drosophila melanogaster*. *Gene Regulation and Systems Biology 2012:6 15-42*

## ABSTRACTS SUBMISSIONS /POSTER PRESENTATIONS

2011 **Kharazmi, J.**, Moshfegh, C., Brody, T. *dmyc* gene is likely to be transcribed form multiple initiation sites. Annual *Drosophila* Meeting, March 30 – April 3, San Diego, Ca.

2010 **Kharazmi, J.** Transcriptional regulation of the *dmyc* gene. Annual *Drosophila* Research Conference, April 7 -11, Washington D.C.

2009 **Kharazmi, J.** Characterization of cis-regulatory elements in the *dmyc* gene of *Drosophila*. Annual *Drosophila* March 4 – 8, Research Conference, Chicago

2008 **Kharazmi, J.** Investigating *dmyc* cis-regulatory elements by studying reporter activity. Annual *Drosophila* Research Conference, April 2 – 6, San Diego, Ca.

2007 **Kharazmi, J.** Phylogenetic Foot Printing of *cis*-Acting Regulatory Elements in the *dmyc* Gene of *Drosophila* Sibling Species. Annual *Drosophila* Research Conference, March 7 – 11, Philadelphia.

2006 **Kharazmi, J.** Site Specific Phage ΦC31 Transgenesis System versus Random P-element Insertion for Studying Reporter Gene Activity. Annual *Drosophila* Research Conference, March 29 – April 2, Huston, Texas.

2005 **Kharazmi, J.**, Moshfegh, C. Investigation of *dmyc* Proximal Regulatory Region. Annual *Drosophila* Research Conference, March 30 – April 3, San Diego, Ca.

2000 **Kharazmi, J.**, (Oral Presentation), Expression of *dmyc* in *Drosophila* Developing Tissues, Swiss Cell Cycle Meeting, Lausanne, Switzerland

## ACQUIRED DIVERSE SKILLS

**Laboratory**

**Microscopy:** Light and Fluorescence: Well-trained, Confocal Microscopy: Basic knowledge

**Profound Knowledge in Working with Model Organism Fruit Fly** Including Genetics, Transformation, Embryo Microinjections, Tissue Preparation and Analyses, Fly Husbandry, Epistasis and Clonal Analyses, Random P-Element and Phage Ci31-Mediated Transgenesis

**Molecular Cloning:** Plasmid Engineering, Polymerase Chain Reaction, Ligation, Plasmid DNA preparation, Electro-Elution, Nucleic Acid Electrophoresis, Dephosphorylation and Phosphorylation of Fragment Ends, In Vitro Transcription and Translation, Nucleic Acid and Protein Isolation from Tissue and Cell Cultures, RNA Modification at the Fragment Ends

**Working with Cell Cultures**

**Flow Cytometry, Fluorescence Activated Cell Sorting (FACS)**

**Nuclear Extraction of Transcription Factors**

**Immuno-Systems** Including Enzyme Linked Immunoabsorbent Assay (ELISA), Immunohistochemistry, Immunoblotting, Immuno-Cytology

**Western Blot and Northern Blot** including PAGE Denaturing Vertical Electrophoresis

**RNA in situ Hybridization on Tissue Sections**

**Revers Transcription and Polymerase Chain Reaction (RT-PCR)**

**5'-Race Amplification of cDNA Ends**

**Preparation of Electro-Competent E.coli Cells**

**Handling and Processing BAC Clone Cosmids**

**DIVERSE COACHING EXPERIENCE**

2012 – 2013 **Mentoring** Hadi Afsari, BBA-Student at Swiss Business School Zurich (SBS)

2007 – 2008 **Instructing** Cameron Moshfegh, on his M.S. Thesis, Molecular Biology, Swiss Federal Institute of Technology Zurich, Laboratory of Biosensors/Bioelectronics Prof. Vörös

**IT SKILLS** **DNASTAR Lasergene 9.1 Core Suit** Well-Trained at Using Modules Primer Select, Seqman, EditSeq and MegAlign

Use of Software programs **Adobe Photoshop CS3, Adobe Illustrator CS5, Microsoft Word and Power Point 2010**

**LANGUAGES** **English,** Excellent **German,** Excellent

**French,** Beginners Level **Persian (Farsi),** Native Language